MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS 1963 A

AD-A179 145

UMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION Unclassified	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release; distribution unlimited
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE APR 16 1987	

4. PERFORMING ORGANIZATION REPORT NUMBER(S)	5. MONITORING ORGANIZATION REPORT NUMBER(S) AFOSR-TR- 87-0318
6a. NAME OF PERFORMING ORGANIZATION University of North Carolina	6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION AFOSR/NM
6c. ADDRESS (City, State and ZIP Code) Department of Statistics Chapel Hill, NC	7b. ADDRESS (City, State and ZIP Code) Bldg 410 Bolling AFB DC 20332-6448
8a. NAME OF FUNDING/SPONSORING ORGANIZATION AFOSR	8b. OFFICE SYMBOL (If applicable) NM	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER F49620-85-C-0144

11. TITLE (Include Security Classification)
Extremal theory for stochastic processes

Professor M.R. Leadbetter

8c. ADDRESS (City, State and ZIP Code) Bolling AFB, DC 20332-6448	10. SOURCE OF FUNDING NOS.

10. SOURCE OF FUNDING NOS.			
PROGRAM ELEMENT NO. 61102F	PROJECT NO. 2304	TASK NO. A5	WORK UNIT NO.

13a. TYPE OF REPORT Reprint	13b. TIME COVERED FROM _____ TO _____	14. DATE OF REPORT (Yr., Mo., Day) November 1985	15. PAGE COUNT 85

16. SUPPLEMENTARY NOTATION

17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB. GR.	

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

The purpose of this paper is to provide an overview of the asymptotic distributional theory of extreme values for a wide class of dependent stochastic sequences and continous parameter processes. The theory contains the standard classical extreme vale results for maxima and extreme order statistics as special cases but is richer on account of the diverse behavior possible under dependence in both discrete and contiuous time contexts. Emphasis is placed on stationairy cases but other important classes (e.g. Markof sequences) are included. Significiant ideas and mithods are described rather than details, and in particular the natureand role of important underlying point processes (such as exceedances and upcrossings) are emphasized. Applications are given to particular classes of process (e.g. normal, moving average) and connections with related theory (such as convergence of sums) are indicated.

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT. ☐ DTIC USERS ☐	21. ABSTRACT SECURITY CLASSIFICATION Unclassified
22a. NAME OF RESPONSIBLE INDIVIDUAL Major Woodruff	22b. TELEPHONE NUMBER (Include Area Code) (202) 767-5026	22c. OFFICE SYMBOL NM

DD FORM 1473, 83 APR EDITION OF 1 JAN 73 IS OBSOLETE.

SECURITY CLASSIFICATION OF THIS PAGE

CENTER FOR STOCHASTIC PROCESSES

Department of Statistics
University of North Carolina
Chapel Hill, North Carolina
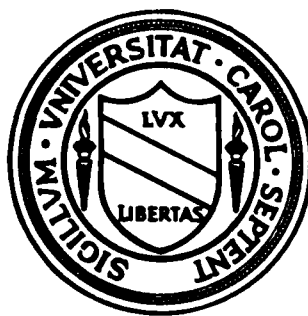
EXTREMAL THEORY FOR STOCHASTIC PROCESSES

by

M.R. Leadbetter

and

H. Rootzén

87

EXTREMAL THEORY FOR STOCHASTIC PROCESSES

by

M.R. Leadbetter   Holger Rootzen
University of North Carolina   University of Copenhagen

The purpose of this paper is to provide an overview of the asymptotic distributional theory of extreme values for a wide class of dependent stochastic sequences and continuous parameter processes. The theory contains the standard classical extreme value results for maxima and extreme order statistics as special cases but is richer on account of the diverse behavior possible under dependence in both discrete and continuous time contexts. Emphasis is placed on stationary cases but other important classes (e.g. Markov sequences) are included. Significant ideas and methods are described rather than details, and in particular the nature and role of important underlying point processes (such as exceedances and upcrossings) are emphasized. Applications are given to particular classes of processes (e.g. normal, moving average) and connections with related theory (such as convergence of sums) are indicated.

CONTENTS

1.1 Scope and content of the paper

The purpose of this paper is to give a "motivated overview" of the principal results in and related to the distributional theory of extremes of stationary sequences and processes. In particular we shall be concerned with distributional properties of the maximum $M_n = \max(\xi_1, \xi_2, \ldots, \xi_n)$ and other order statistics from stationary sequences $\{\xi_i\}$ as $n \to \infty$ and with corresponding results for continuous parameter processes. The emphasis throughout will be on the motivation for and significant methods used in obtaining the results. Full proofs will not generally be given - in many cases the details of such proofs may be found in the volume [55], or from the references cited.

The results to be described may, in part, be regarded as extensions of the classical theory of extremes of sequences of independent, identically distributed (i.i.d.) random variables (r.v.'s). However, they constitute more than just such an extension of the classical theory, since the dependent framework provides a natural setting for the theory and one in which its essential ideas and methods may be clearly exposed. In particular, it will be seen that the central results may often be regarded as special cases of the convergence of certain point processes - a view which may of course be taken in the classical case but which is less needed there in view of the detailed i.i.d. assumptions. Our discussion will emphasize the centrality of these underlying point process convergence results.

As indicated in the list of contents, this paper is organized in three main parts. This first introductory part contains central distributional results of the classical i.i.d. theory and, in particular, the "Extremal Types Theorem" which restricts the possible limiting distributions for maxima to essentially three "different types". We shall indicate the general

A-1

organization and main features of the most recently available derivations of these results, using the simple and elegant approach of de Haan via inverse functions. As can be seen even this i.i.d. theory becomes most natural and transparent when viewed from the standpoint of the behavior of related point processes – such as the exceedances of high levels.

The second part of the paper concerns extremes of sequences – primarily (but not always) assumed stationary and is largely based on point process methods. It will be seen that the classical theory may be regarded as a special case of the more general theory for dependent sequences – some results being identical and others generalizing in interesting and non-trivial ways. For example, it will be seen, under weak dependence restrictions, that the general "type" of limiting distribution for the maximum is the same as for an i.i.d. sequence with the same marginal d.f. (though the normalizing constants may change). However, the limiting distributions for other order statistics can be quite different from those under i.i.d. assumptions.

Some particular cases of special interest (e.g. normal sequences, moving averages, Markov sequences) will be discussed in Part 2. Other aspects of the theory (e.g. rates of convergence, multivariate extremes) are also briefly described along with some interesting connections with convergence of sums.

In Part 3 attention is turned to continuous parameter processes. The theory here may be made to rest on the sequence case by the simple device of regarding the maximum of a process $\xi(t)$ up to, say, time $T = n$ as the maximum of the values of the sequence $\zeta_i = \sup\{\xi(t): i-1 \le t \le i\}$, for $1 \le i \le n$. While this is simple and obvious in principle, the details are more complicated and require analogous but somewhat more intricate assumptions regarding the dependence structure of the process. The point process approach is also very

valuable here - considering, for example, upcrossings of high levels in lieu of exceedances. Again, a rather full and satisfying theory results and is applied, in particular, to special cases such as normal, and χ^2 processes. Properties of point processes of *local maxima* may also be obtained, as will be briefly indicated.

It may be noted that the *stationarity* assumption, where made, primarily provides for convenience and clarity, and that some departures from this will either not alter the result, or will alter it in an interesting way which can be determined. This will be evident, e.g. in discussion of normal sequences, where extensions to useful non-stationary cases will be briefly mentioned. Finally this paper is not by any means intended as a complete review of all aspects of extremal theory - a number of important topics are not referred to at all. Rather it is our purpose to provide an overview of much of a developing area which includes but is more widely applicable than the classical theory, and is based on the interplay of interesting mathematical techniques. In particular we emphasize recent results - especially those obtained since the publication of the volume [55].

1.2 Classical extreme value theory

The principal concern of classical extreme value theory is with asymptotic distributional properties of the maximum $M_n = \max(\xi_1, \xi_2, \ldots, \xi_n)$ from an i.i.d. sequence $\{\xi_i\}$ as $n \to \infty$. While the distribution function (d.f.) of M_n may be written down exactly ($P\{M_n \leq x\} = F^n(x)$ where F is the d.f. of each ξ_i), there is nevertheless virtue in obtaining asymptotic distributions which are less dependent on the precise form of F, i.e. relations of the form

$$(1.2.1) \qquad P\{a_n(M_n - b_n) \leq x\} \xrightarrow{d} G(x) \qquad \text{as } n \to \infty,$$

where G is a non-degenerate d.f. and $a_n > 0$, b_n, are normalizing constants.

The central result of classical extreme value theory, due in varying degrees of generality to Frechet [35], Fisher and Tippett [34], and Gnedenko [37], restricts the class of possible limiting d.f's G in (1.2.1) to essentially three different types as follows.

<u>Theorem</u> 1.2.1 <u>(Extremal Types Theorem)</u>. Let $M_n = \max(\xi_1, \xi_2, \ldots, \xi_n)$ where ξ_i are i.i.d. If (1.2.1) holds for some constants $a_n > 0$, b_n and some non-degenerate G, then G must have one of the following forms (in which x may be replaced by $ax + b$ for any $a > 0$, b):

$$\text{Type I:}\quad G(x) = \exp(-e^{-x}) \qquad -\infty < x < \infty$$

$$\text{Type II:}\quad G(x) = \begin{cases} 0 & x \leq 0, \\ \exp(-x^{-\alpha}), \text{ for some } \alpha > 0 & x > 0 \end{cases}$$

$$\text{Type III:}\quad G(x) = \begin{cases} \exp(-(-x)^{\alpha}), \quad \text{for some } \alpha > 0 & x \leq 0, \\ 1, & x > 0. \end{cases}$$

Conversely any such d.f. G may appear as a limit in (1.2.1) and in fact does so when G is itself the d.f. of each ξ_i.

It will be convenient to say that two non-degenerate d.f.'s G_1 and G_2 are of the *same type* if $G_1(x) = G_2(ax + b)$ for some $a > 0$, b, and to refer to the equivalence classes so determined as "types". The use of "type" in the above theorem is a slight abuse of this since Types II and III really represent *families* of types—one corresponding to each $\alpha > 0$. However this abuse is convenient and it is conventional to refer to "the three types" of limit. It should also be noted that the three types may be incorporated into a single family, for example by writing $G_\alpha(x) = \exp\{-(1-\alpha x)^{1/\alpha}\}$, $-\infty < \alpha < \infty$, $\alpha x < 1$, G_0 being interpreted as $\lim_{\alpha \to 0} G_\alpha(x) = \exp(-e^{-x})$. (Such a parametrization was introduced by von Mises).

Before indicating the main features of the proof of this theorem, it

will be convenient to state the widely used result of Khintchine which enters
extremal theory here and elsewhere in significant ways.

Lemma 1.2.2 (Khintchine). Let $\{F_n\}$ be a sequence of d.f.'s and G a non-
degenerate d.f. Let $a_n > 0$, b_n be constants such that $F_n(a_n x + b_n) \overset{d}{\to} G(x)$.
Then for some nondegenerate d.f. G_* and constants $\alpha_n > 0, \beta_n$,
$F_n(\alpha_n x + \beta_n) \overset{d}{\to} G_*(x)$ if and only if $a_n^{-1}\alpha_n \to a$ and $a_n^{-1}(\beta_n - b_n) \to b$ for some
$a > 0$ and b, and then $G_*(x) = G(ax+b)$ so that G and G_* are of the same type.

The derivation of Theorem 1.2.1 can be conveniently divided into two
parts and this division is most useful for later forms of the result. The
first part is to show that the class of limit laws G in (1.2.1) is precisely
the class of *max-stable* d.f.'s. Specifically a d.f. G is called *max-
stable* if for each n=1,2,... the d.f. G^n is of the same type as G, i.e. if
there exist constants $a_n > 0, b_n$ such that $G^n(a_n x + b_n) = G(x)$. The following
lemma—which follows readily using Lemma 1.2.2—gives an immediate proof of
this part.

Lemma 1.2.3. A non-degenerate d.f. G is max-stable if and only if there is a
sequence $\{F_n\}$ of d.f.'s, and constants $a_n > 0, b_n$ such that

(1.2.2) $\quad F_n(a_{nk}^{-1}x + b_{nk}) \overset{d}{\to} G^{1/k}(x)$ as $n \to \infty$ for each k=1,2...

It is easily seen from this that any G satisfying (1.2.1) is max-
stable by simply identifying F_n with F^n where F is the d.f. of each ξ_i. For
(1.2.1) is just $F^n(a_n^{-1}x+b_n) \overset{d}{\to} G(x)$ and replacing n by nk yields (1.2.2) at
once. The converse is even simpler since if G is max stable and
$G^n(a_n^{-1}x+b_n) = G(x)$ for some $a_n > 0$, b_n, an i.i.d. sequence with common d.f. G
satisfies

$$P\{a_n(M_n - b_n) \le x\} = G^n(a_n^{-1}x + b_n) = G(x)$$

so that (1.2.1) holds trivially and G is indeed a limit law for maxima.

The essential point of this argument is that if F_n is the d.f. of M_n, then the assumption (1.2.1) is the same as (1.2.2) with k = 1 and the independence shows that if (1.2.2) holds with k = 1, it holds with k = 2,3,... so that G is max-stable. It will be seen later that the same line of argument holds in *dependent* cases though the details are more complicated since F_n is no longer precisely F^n.

Thus the class of non-degenerate limit laws for maxima is precisely the class of max-stable d.f.'s. The other part to the proof of Theorem 1.2.1 is to identify the class of max-stable d.f.'s with the Type I, II and III extreme value d.f.'s. This is a purely function-analytic (non-probabilistic) procedure and will apply verbatim in dependent cases. It is in fact trivial to show that each extreme value d.f. is max-stable. The more important converse is readily shown by transforming the definition of max-stability of G to give a functional equation for the function U(y) inverse to $-\log(-\log G(x))$, which may be readily solved (cf. [38]) to show that G must be one of the three extreme value types.

It is, of course, important to know which (if any) of the three types of limit law applies when ξ_n has a given d.f. F. Necessary and sufficient conditions are known, involving only the behavior of the tail 1-F(x) as x increases for each possible limit. One form of such domain of attraction criteria is as follows. In this x_F ($\le \infty$) will denote the right hand end point sup(t: F(t) < 1) of the d.f. F.

Theorem 1.2.4. Necessary and sufficient conditions for the d.f. F of the r.v.'s of the i.i.d. sequence $\{\xi_n\}$ to lead to each of the three types are:

Type I: There exists some strictly positive function g(t) such that

$$\lim_{t \uparrow x_F} (1-F(t+xg(t)))/(1-F(t)) = e^{-x} \quad \text{for all real } x;$$

Type II: $x_F = \infty$ and

$$\lim_{t \to \infty} (1-F(tx))/(1-F(t)) = x^{-\alpha}, \; \alpha > 0, \text{ for each } x > 0;$$

Type III: $x_F < \infty$ and

$$\lim_{h \downarrow 0} (1-F(x_F-xh))/(1-F(x_F-h)) = x^{\alpha}, \; \alpha>0, \text{ for each } x > 0.$$

The sufficiency of these conditions is readily established (cf. [55]). The necessity is more complicated (though perhaps also less important) but may be achieved by using methods of regular variation (cf. [39] for a recent smooth treatment).

The following simple result is also used in these "domain of attraction" determinations.

<u>Theorem</u> <u>1.2.5.</u> Let $\{u_n, \; n \geq 1\}$ be constants and $0 \leq \tau \leq \infty$. If $\xi_1, \xi_2, \ldots$ are i.i.d. with d.f. F then

(1.2.3) $\qquad P\{M_n \leq u_n\} \to e^{-\tau}$

if and only if

(1.2.4) $\qquad n(1 - F(u_n)) \to \tau.$

This result is proved almost trivially by writing $P\{M_n \leq u_n\} = F^n(u_n)$ $= (1-(1-F(u_n)))^n$ in this i.i.d. context. It is stated formally since its generalizations to dependent cases are important and much less trivial in proof. It may be noted that (1.2.1) is a special case of (1.2.3) using a linear parametrization, by making identifications $\tau = -\log G(x)$, $u_n = a_n^{-1}x + b_n$. Thus a necessary and sufficient condition for the limit G is

$$n(1 - F(a_n^{-1}x + b_n)) \to -\log G(x), \quad \text{as } n \to \infty,$$

for each x, and some $a_n > 0$, b_n. This explains the relevance of the tail

$1 - F(x)$ in Theorem 1.2.4 which examines the existence of such a_n, b_n for each of the cases $-\log G(x) = e^{-x}$, $x^{-\alpha}$, $(-x)^{\alpha}$.

By applying this result, forms for the normalizing constants may also be obtained. Specifically if γ_n is defined to satisfy $F(\gamma_n -) \leq 1-n^{-1} \leq F(\gamma_n)$, a_n and b_n may be taken in each case to be:

	a_n	b_n
Type I:	$(g(\gamma_n))^{-1}$	γ_n
Type II:	γ_n^{-1}	0
Type III:	$(x_F - \gamma_n)^{-1}$	x_F

(using the notation of Theorem 1.2.4). Of course while the "$(1-n^{-1})$-percentile γ_n may be determined (and hence a_n, b_n found) when F is known, the practical problem lies in the estimation of those constants when the form of F is not precisely known.

It is readily checked that a standard normal sequence belongs to the Type I domain with normalizing constants

$$(1.2.5) \qquad a_n = (2 \log n)^{1/2}$$
$$b_n = (2 \log n)^{1/2} - \frac{1}{2} (2 \log n)^{-1/2}(\log \log n + \log 4\pi).$$

The exponential and log normal distributions also have Type I limits as does the d.f. $F(x) = 1-e^{1/x}$ $(x < 0)$ with a finite right endpoint $x_F = 0$.

The Pareto and Cauchy distributions give Type II limits whereas the uniform distribution belongs to the Type III domain.

It should be noted that not every d.f. F belongs to a domain of attraction at all. The most common case occurs for certain discrete distributions--such as the Poisson and geometric distributions--for which there is no sequence $\{u_n\}$ such that (1.2.4) holds. This typically happens in cases when the jumps of the d.f. do not decay sufficiently quickly relative to

the tail. In fact for a given τ $(0<\tau<\infty)$, a sequence $\{u_n\}$ satisfying (1.2.4) may be found if and only if

(1.2.6) $(F(x)-F(x-))/(1-F(x)) \to 0$ as $x \uparrow x_F$.

It is readily checked that (1.2.6) fails for the Poisson and geometric cases and hence there can be no u_n of any form (and certainly not of the form $a_n^{-1}x + b_n$) satisfying (1.2.4) with $\tau = -\log G(x)$, so that no limiting distribution exists. However it is also possible for there to be no limit even if (1.2.6) is satisfied—indeed for certain continuous d.f.'s. A case in point is the d.f. $F(x) = 1-e^{-x-\sin x}$, an example due to Von Mises.

We turn now, in this brief tour of classical results, to other extreme order statistics, writing $M_n^{(k)}$ for the kth largest among the i.i.d. $\xi_1,\dots,\xi_n$ with common d.f. F. Suppose that (1.2.6) holds and hence for any fixed $\tau>0$, $u_n = u_n(\tau)$ may be found satisfying (1.2.4). Let S_n denote the number of *exceedances* of u_n by $\xi_1,\dots,\xi_n$, i.e. S_n is the number of i, $1\le i\le n$, such that $\xi_i > u_n$. Clearly for $k = 1,2,\dots,$

(1.2.7) $P\{M_n^{(k)} \le u_n\} = P\{S_n < k\}$

since the events in brackets are identical. But S_n is binomial with parameters (n,p_n), $p_n = 1 - F(u_n)$, $np_n \to \tau$ so that S_n has a Poisson limit with mean τ and hence

(1.2.8) $P\{M_n^{(k)} \le u_n\} \to e^{-\tau} \sum_{r=0}^{k-1} \tau^r/r!$

Suppose now that $M_n = \max(\xi_1,\xi_2,\dots,\xi_n)$ has limiting distribution G so that (1.2.1) holds. By the standard identification $u_n = a_n^{-1}x+b_n$, $\tau = -\log G(x)$ we see from Theorem 1.2.5 that (1.2.4) holds and hence from (1.2.8) that

(1.2.9) $P\{a_n(M_n^{(k)} - b_n) \le x\} \to G(x) \sum_{s=0}^{k-1} (-\log G(x))^s/s!$

Thus if the maximum M_n has a limiting distribution G, then the kth largest $M_n^{(k)}$ has a limiting distribution given by (1.2.9) (with the *same* normalizing constants a_n, b_n as the maximum itself).

These results foreshadow a more detailed discussion of the exceedances and related point processes, which will be taken up in the next section.

Finally, topics from the classical theory not dealt with in this present part include (a) rate of convergence results (considered in the dependent setting in Section 2.8, (b) asymptotic distributions of minima (obtainable by simple transformations of the results for maxima), and (c) asymptotic theory of variable rank order statistics (cf. [85]).

1.3 Point processes associated with extremes

The heart of the previous simple calculation leading to (1.2.9) is that if $n(1-F(u_n)) \to \tau$ then the number of exceedances of u_n by $\xi_1 \ldots \xi_n$ is asymptotically Poisson with mean τ. This simple observation is capable of considerable, useful generalization both for the present i.i.d. and for dependent cases. The simplest of these results concerns the *point process* N_n of *exceedances* of the level u_n. Specifically N_n consists of the point process on $(0,1]$ formed by normalizing the actual exceedance points by the factor $1/n$ i.e. if i is the time of an exceedance ($\xi_i > u_n$) then a point of N_n is plotted at i/n. If $E \subset (0,1]$ then $N_n(E)$ denotes the number of such points in E, so that $N_n(E) = \#\{i/n \in E: \xi_i > u_n, 1 \le i \le n\} = \#\{i \in nE: \xi_i > u_n, 1 \le i \le n\}$. The actual exceedance points and the point process N_n are illustrated in Figure 1.3.1 below.
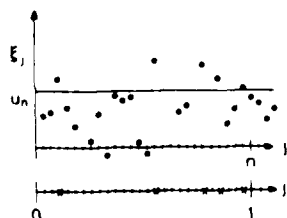

Figure 1.3.1 Point process of exceedances

One of the central results is that the point process N_n takes on an asymptotic Poisson character as n increases in the sense that N_n converges in distribution to a Poisson process N as $n \to \infty$ (Theorem 1.3.1 below). A brief discussion of relevant features of point processes and their properties will be given in Section 2.3 but for now it is sufficient to note that convergence in distribution of point processes may be expressed in terms of convergence of finite dimensional distributions; specifically $N_n \overset{d}{\to} N$ if $N(\{a\})=0$ for each a and

(1.3.1) $(N_n(a_1,b_1], N_n(a_2,b_2]...N_n(a_k,b_k]) \overset{d}{\to} (N(a_1,b_1], N(a_2,b_2],...N(a_k,b_k])$

for each choice of k and subintervals $(a_i,b_i] \subset (0,1]$.

<u>Theorem 1.3.1</u> Let $\xi_1, \xi_2...$ be i.i.d. with common d.f. F and let $\{u_n\}$ satisfy (1.2.4). Then $N_n \overset{d}{\to} N$, where N is a Poisson process on (0,1] with intensity τ.

This result is almost selfevident by virtue of the criterion (1.3.1). For example $N_n((a,b])$ is the number of $i \in (na,nb]$ for which $\xi_i > u_n$ and hence is binomial with parameters $([nb]-[na]), 1-F(u_n))$ and converges to the Poisson r.v. $N((a,b])$ having mean $\tau(b-a)$ since $([nb]-[na])(1-F(u_n)) \to \tau(b-a)$ by (1.2.4). The more general statement (1.3.1) clearly follows by independence if the intervals $(a_i,b_i]$ are disjoint, and in general by considering the overlaps between the intervals and thereby reducing the problem to the case of disjoint intervals in an obvious way.

Theorem 1.3.1 clearly includes (1.2.7) and hence (1.2.9) since $N_n((0,1])=S_n$ so that it may be regarded as a "fountainhead" result from which the asymptotic distributions for the maximum and all extreme order statistics follow. The result may be extended by considering more than one leve? to give in particular asymptotic joint distributions of finite numbers of order statistics. Specifically let $0 < \tau_1 < \tau_2...\le \tau_r$ be fixed constants and $\{u_n(\tau)\}$ such that

(1.3.2)
$$n(1 - F(u_n(\tau))) \to \tau$$

for each $\tau > 0$, where $u_n(\tau)$ are taken so that $u_n(\tau_1) \geq u_n(\tau_2)\dots \geq u_n(\tau_r)$.
Consider exceedances of the levels $u_n(\tau_1),\dots,u_n(\tau_r)$ as illustrated in Fig.
1.3.2(a) (again normalizing the time scale by $1/n$). A vector point process is
thus obtained and may be visualized by points on fixed lines $L_1, L_2,\dots, L_r$ in

the plane as shown in Fig. 1.3.2(b). Denoting these individual point processes by
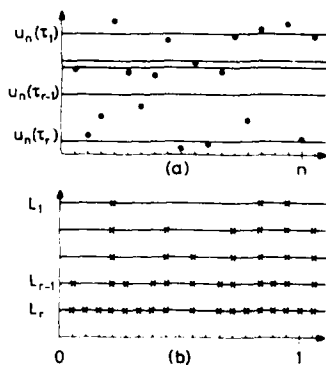
Figure 1.3.2 (a) Levels and values of $\{\xi_n\}$ (b) Representation on fixed lines
$L_1, L_2,\dots, L_r.$ $L_1, L_2,\dots, L_s$

$N_n^{(k)}$, $1 \leq k \leq r$, it is clear that each $N_n^{(k)}$ is asymptotically Poisson with
intensity τ_k, and that $N_n^{(k-1)}$ is a "thinned" version of $N_n^{(k)}$ for $2 \leq k \leq r$.
In fact in the limit the thinnings involve independent removal of events.
To see this more explicitly let N_n be the process defined on the plane by
the points of $N_n^{(k)}$ $1 \leq k \leq r$ (i.e. confined to the lines $L_1, L_2,\dots, L_r$) and
let N be a point process in the plane defined as follows:

Let $\{\sigma_{1j};\ j=1,2,\dots\}$ be the points of a Poisson Process $N^{(r)}$ with
parameter τ_r on L_r and let $\{\beta_j,\ j=1,2,\dots\}$ be i.i.d. random variables,
independent also of the Poisson process on L_r, taking values $1,2,\dots,r$ with
probabilities

$$P\{\beta_j = s\} = \begin{cases} (\tau_{r-s+1} - \tau_{r-s})/\tau_r\ , & s=1,2,\dots,r-1, \\ \tau_1/t_r\ , & s=r, \end{cases}$$

i.e. $P\{\beta_j > s\}= \tau_{r-s+1}/\tau_r$ for $s=1,2,\ldots,r$. For each j, place points

$\sigma_{2j}, \sigma_{3j}, \ldots, \sigma_{\beta_j j}$ on the β_j-1 lines $L_{r-1}, L_{r-2}, \ldots, L_{r-\beta_j+1}$, vertically

above σ_{1j}, to complete the point process N. Clearly the probability that a

point appears on L_{r-1} above σ_{1j} is just $P\{\beta_j \geq 2\}=\tau_{r-1}/\tau_r$ and the deletions are

independent, so that $N^{(r-1)}$ is obtained on L_{r-1} as an independent thinning of

the Poisson process $N^{(r)}$. Hence $N^{(r-1)}$ is a Poisson process with intensity

$\tau_r(\tau_{r-1}/\tau_r)=\tau_{r-1}$ as expected. Similarly $N^{(k)}$ is obtained on L_k as an

independent thinning of $N^{(k+1)}$ with deletion probability $1-\tau_k/\tau_{k+1}$, all $N^{(k)}$

being Poisson. The main theorem is as follows.

<u>Theorem 1.3.2</u>. The point processes N_n defined in the plane as above (with

levels satisfying (1.3.2)) converge in distribution as $n \to \infty$ to the point

process N consisting of r successively thinned Poisson processes.

Again the proof of this may be accomplished by showing that

$(N_n(B_1),\ldots,N_n(B_k)) \overset{d}{\to} (N(B_1),\ldots,N(B_k))$ for each choice of k and rectangles

$B_1,\ldots,B_k$, the calculations being more involved but similar to the one-

dimensional case.

Theorem 1.3.2 may be used to give the asymptotic joint distribution of

extreme order statistics (and their locations). For example the following

result concerns the maximum and second largest values.

<u>Theorem 1.3.3</u> Suppose that the maximum $M_n(=M_n^{(1)})$ has the asymptotic distribution G,

(1.3.3) $\qquad P\{a_n(M_n^{(1)}-b_n) \leq x\} \to G(x) \qquad$ as $n \to \infty$.

Then for $x_1 > x_2$,

(1.3.4) $\quad P\{a_n(M_n^{(1)}-b_n) \leq x_1, \ a_n(M_n^{(2)}-b_n) \leq x_2\} \to$

$$G(x_2)(\log G(x_1) - \log G(x_2) + 1) \qquad \text{as } n \to \infty.$$

This may be proved by writing $\tau_i = -\log G(x_i)$, $u_n(\tau_i) = x_i/a_n+b_n$,

i=1,2 and noting that (1.3.3) implies (using Theorem 1.2.5) that $n(1-F(u_n(\tau_i))) \to \tau_i$ i=1,2. The left hand side of (1.3.4) is then just (writing $N_n^{(k)}$ for $N_n^{(k)}((0,1])$),

$$P\{M_n^{(1)} \le u_n(\tau_1), M_n^{(2)} \le u_n(\tau_2)\} = P\{N_n^{(2)}=0\} + P\{N_n^{(1)}=0, N_n^{(2)}=1\}$$

which converges to the same probabilities with $N^{(k)}$ replacing $N_n^{(k)}$. This is readily seen to be $e^{-\tau_2} + (\tau_2-\tau_1)e^{-\tau_2}$ which is just the right hand side of (1.3.4).

The "r-level" result Theorem 1.3.2 allows the joint asymptotic distribution of r extreme order statistics to be obtained. On the other hand these results may be summarized in one theorem commonly referred to as a "complete" convergence result, and which concerns convergence of the process values themselves (suitably normalized) regarded as a point process in the plane. This result is intuitively satisfying and in the i.i.d. case it may be regarded as the fundamental result yielding all the relevant asymptotic distributional properties. On the other hand when dependence is introduced into a sequence the "partial" r-level results require somewhat less assumptions than does the "complete" result.

<u>Theorem</u> <u>1.3.4</u> Let $\xi_1, \xi_2 \ldots$ be i.i.d. with marginal d.f. F. Suppose that $u_n(\tau)$, satisfying (1.3.2) for each $\tau>0$, is continuous and strictly decreasing in τ for each n. Let N_n^* denote the point process in the plane consisting of the points $(j/n, u_n^{-1}(\xi_j))$ where u_n^{-1} denotes the inverse function of $u_n(\tau)$, defined on the range of the r.v.'s $\{\xi_j\}$. Then $N_n^* \overset{d}{\to} N^*$, where N^* is a Poisson process on $(0,\infty) \times (0,\infty)$ having Lebesgue measure m as its intensity.

<u>Proof</u>. Again this is readily proved from convergence of finite dimensional

distributions. For a rectangle $B=(a,b]\times[c,d)$, $N_n^*(B)$ is clearly binomial with parameters $([nb]-[na],p_n)$ where $p_n=P\{c \le u_n^{-1}(\xi_1) < d\} = F(u_n(c))-F(u_n(d))$ $(d-c)/n$, so that $([nb]-[na])p_n \to (b-a)(d-c) = m(B)$. Hence $N_n^*(B)$ converges in distribution to a Poisson r.v. with mean $m(B)$. If $B_i=(a_i,b_i]\times\bigcup_j[c_{ij},d_{ij})$ where the intervals $(a_ib_i]$ are disjoint, it follows along similar lines that $(N_n^*(B_1)\ldots N_n^*(B_k))) \overset{d}{\to} (N^*(B_1)\ldots N^*(B_k)))$. It is readily seen from this that the same holds when B_i are arbitrary rectangles of the form $(a_i,b_i]\times[c_i,d_i)$ which is sufficient (cf. Sec. 2.3) to show that $N_n^* \overset{d}{\to} N^*$.

As noted above this theorem summarizes a whole spectrum of asymptotic distributional results for maxima and extreme order statistics in the i.i.d. case. For example $P\{N_n((0,1]\times[0,\tau)) < r\} = P\{M_n^{(r)} \le u_n(\tau)\}$ as is easily checked so that the limiting distribution of $M_n^{(r)}$ can be obtained. When M_n has the asymptotic distribution G as in (1.2.1) we may take $u_n(\tau) = a_n^{-1}G^{-1}(e^{-\tau})+b_n$ as is readily checked. In that case Theorem 1.3.4 may be readily transformed to give the following form (writing $x_0 = \inf\{x: G(x) > 0\}$).

Theorem 1.3.5 Suppose (1.2.1) holds, for the i.i.d. sequence $\{\xi_j\}$, and let N_n' be the point process in the plane with points at $(j/n, a_n(\xi_j-b_n))$. Then $N_n' \to N'$ on $(0,\infty) \times (x_0,\infty)$ where N' is a Poisson process whose intensity measure is the product of Lebesgue measure and that defined by the increasing function $\log G(y)$.

This form of the result was first proved by Pickands ([71]) and is more transparent when linear normalizations give an asymptotic distribution for M_n. (Theorem 1.3.4 applies to linear or nonlinear normalizations). For example it is clear that $P\{a_n(M_n^{(r)}-b_n) \le x\} = P\{N_n'((0,1] \times (x,\infty)) \le r-1\} \to P\{N'((0,1] \times (x,\infty)) \le r-1\}$ from which (1.2.9) follows simply.

2. Extremes of sequences

2.1 The Extremal Types Theorem for stationary sequences.

In this section it will be shown that the Extremal Types Theorem still holds for (strictly) stationary sequences under weak dependence assumptions. Obviously some form of restriction on the dependence structure of the sequence is necessary to obtain nontrivial results since e.g. one might take all ξ_i to be equal with arbitrary d.f., so that M_n would also have this assigned d.f. Then in the next section we shall see that the introduction of dependence does not typically alter the limiting distributional type for the maximum and will explore the precise changes involved.

Loynes ([61]) first obtained a form of the Extremal Types Theorem under dependence - assuming strong mixing. Weaker (distributional) conditions will suffice and will be used here. The difference is not too important for our present purposes since the main ideas of proof are essentially the same. The main condition to be used (termed $D(u_n)$) is defined with reference to a sequence $\{u_n\}$ of constants in terms of the finite dimensional d.f.'s

$$F_{i_1 \ldots i_n}(x_1, \ldots x_n) = P\{\xi_{i_1} \leq x_1, \ldots \xi_{i_n} \leq x_n\}$$ of the stationary sequence $\{\xi_n\}$.

Writing $F_{i_1, \ldots i_n}(u) = F_{i_1, \ldots i_n}(u, u, \ldots u)$, define

$$\alpha_{n,l} = \max\{| F_{i_1, \ldots i_p, j_1 \ldots j_p}(u_n) - F_{i_1, \ldots i_p}(u_n) \, F_{j_1, \ldots j_p}(u_n)| :$$

$$1 \leq i_1 < i_2 < i_p < j_1 \ldots < j_p, \leq n, \; j_1 - i_p \geq l\}$$

Then $D(u_n)$ is said to hold if $\alpha_{n,l_n} \to 0$ for some sequence $l_n = o(n)$.

It is, incidentally, obviously possible to weaken the condition $D(u_n)$ very slightly to involve "intervals" of consecutive integers (See O'Brien ([68]) for the details of such a procedure and for some advantages in

application to periodic Markov chains.)

The following result is basic for the discussion of M_n and shows the form in which $D(u_n)$ entails approximate independence. It is stated in a somewhat more general form than needed in this section.

Lemma 2.1.1. Let $\{u_n\}$ be a sequence of constants and let $D(u_n)$ be satisfied by the stationary sequence $\{\xi_n\}$. Let $\{k_n \geq 1\}$ be constants such that $k_n = o(n)$ and (in the notation used above for $D(u_n)$), $k_n l_n = o(n)$, $k_n \alpha_{n,l_n} \to 0$. Then

$$P\{M_n \leq u_n\} - P^{k_n}\{M_{r_n} \leq u_n\} \to 0 \qquad \text{as } n \to \infty,$$

where $r_n = [n/k_n]$.

The proof of this result is perhaps the key method in dependent extremal theory. The type of argument was used first in this context by Loynes ([61]) but was used earlier in dependent central limit theory (cf. [15]).

The basic idea is to divide the integers $1, 2, \ldots, n$ into "intervals" $I_1, I_1^*, I_2, I_2^* \ldots I_k, I_k^*$, where $I_1 = (1, 2 \ldots r_n - l_n)$, $I_1^* = (r_n - l_n + 1, \ldots r_n)$ $I_2 = (r_n + 1, \ldots 2r_n - l_n)$ and so on in this fashion except for the last interval $I_{k_n}^* = (k_n r_n - l_n + 1, \ldots n)$. Thus $I_1, I_2 \ldots I_{k_n}$ are large intervals separated by smaller (but typically expanding) intervals $I_1^*, I_2^* \ldots$. The steps of proof (cf. ([54]) for details) are

    (i) Approximate $P\{M_n \leq u_n\}$ by $P\{\bigcap\limits_{j=1}^{k_n} (M(I_j) \leq u_n)\}$ (using $M(E)$ to denote $\max\{\xi_j : j \in E\}$). This simply reflects the fact that the maximum is likely to occur on the larger intervals.

    (ii) Approximate $P\{\bigcap\limits_{j=1}^{k_n} (M(I_j) \leq u_n)\}$ by $P^{k_n}\{M(I_1) \leq u_n\}$ using $D(u_n)$.

    (iii) Approximate $P^{k_n}\{M(I_1) \leq u_n\}$ by $P^{k_n}\{M_{r_n} \leq u_n\}$, the maximum of the first r_n of the ξ_i's being likely to occur on I_1.

The Extremal Types Theorem now follows simply from this result.

__Theorem__ 2.1.2. (Extremal Types Theorem for Stationary Sequences) Let $\{\xi_n\}$ be a stationary sequence such that $M_n = \max(\xi_1, \xi_2, \ldots, \xi_n)$ has a non-degenerate limiting distribution G as in (1.2.1). Suppose that $D(u_n)$ holds for each u_n of the form $u_n = x/a_n + b_n$, for x with $0 < G(x) < 1$. Then G is one of the three classical extremal types.

__Proof__. Writing $u_n = x/a_n + b_n$ it follows that $P\{M_n \leq u_n\} \to G(x)$ and $D(u_n)$ holds (at continuity points of G). By Lemma 2.1.1 by putting $k_n = k$, fixed, and then replacing n by nk we have $P\{M_{nk} \leq u_{nk}\} - P^k\{M_n \leq u_{nk}\} \to 0$ so that $P\{M_n \leq u_{nk}\} \to G^{1/k}(x)$ or $P\{a_{nk}(M_n - b_{nk}) \leq x\} \to G^{1/k}(x)$, $k = 1, 2, \ldots$. But this implies that G is max stable by Lemma 1.2.3 and hence an extreme value d.f.

2.2 The Extremal Index

While the introduction of dependence into a sequence can significantly affect various extremal properties, it does not, within broad limits, affect the distributional type for the maximum. The purpose of this section is to make that rough statement precise and to explore the explicit changes brought by a dependence structure. This depends essentially on a single parameter sometimes called the "extremal index" of the (stationary) sequence $\{\xi_n\}$.

Following Loynes ([61]) it will be convenient, for a given stationary sequence $\{\xi_n\}$, to define the *associated independent sequence* $\{\hat{\xi}_n\}$ to be i.i.d. with the same d.f. F as ξ_n and to write $\hat{M}_n = \max(\hat{\xi}_1, \hat{\xi}_2, \ldots \hat{\xi}_n)$, with $M_n = \max(\xi_1, \xi_2, \ldots \xi_n)$ as before. As noted originally for strongly mixing sequences in [61], if $u_n = u_n(\tau)$ satisfies (1.2.4) for each τ, i.e.

(2.2.1) $$n[1 - F(u_n(\tau))] \to \tau$$

then any limit (function) for $P(M_n \leq u_n(\tau))$ must be of the form $e^{-\theta \tau}$ with

fixed $\theta \in [0,1]$ rather than just the function $e^{-\tau}$ given by (1.2.3) in the i.i.d. case.

   If $P\{M_n \le u_n(\tau)\} \to e^{-\theta\tau}$ for each $\tau > 0$, with $u_n(\tau)$ satisfying (2.2.1), we say that the stationary sequence $\{\xi_n\}$ *has extremal index* θ (≥ 0). This definition does not involve any dependence restriction on the sequence $\{\xi_n\}$. The following result shows that, under $D(u_n)$ conditions any limit for $P\{M_n \le u_n(\tau)\}$ must be of this form.

Lemma 2.2.1. For the stationary sequence $\{\xi_n\}$ and constants $\{u_n(\tau)\}$ satisfying (2.2.1) suppose that $D(u_n(\tau))$ holds for each $\tau > 0$. Then there exist constants θ, θ', $0 \le \theta \le \theta' \le 1$ such that $\lim \sup_{n \to \infty} P\{M_n \le u_n(\tau)\} = e^{-\theta\tau}$, $\lim \inf_{n \to \infty} P\{M_n \le u_n(\tau)\} = e^{-\theta'\tau}$ for each τ, so that if $P\{M_n \le u_n(\tau)\}$ converges for some $\tau > 0$ then $\theta' = \theta$ and $P\{M_n \le u_n(\tau)\} \to e^{-\theta\tau}$ for all $\tau > 0$ and $\{\xi_n\}$ has extremal index θ, $0 \le \theta \le 1$.

   This result is proved by using Lemma 2.1.1 to show that $\psi(\tau) = \lim \sup P\{M_n \le u_n(\tau)\}$ satisfies $\psi(\tau/k) = \psi^{1/k}(\tau)$ for each $k = 1, 2, \ldots$, to give the exponential limit. The details of this proof may be found in [54]. Note that it follows in the course of the proof that $\theta \le 1$.

   Clearly for any i.i.d. sequence for which (1.2.6) holds (so that $u_n(\tau)$ may be found to satisfy (2.2.1)) has extremal index $\theta = 1$. A stationary sequence $\{\xi_n\}$ satisfying $D(u_n(\tau))$ for each $\tau > 0$ also has extremal index $\theta = 1$ if

(2.2.2) $\lim \sup_{n \to \infty} n \sum_{j=2}^{[n/k]} P\{\xi_1 > u_n, \xi_j > u_n\} \to 0$ as $k \to \infty$.

For proof see [55] where (2.2.2) is referred to as $D'(u_n)$.

   Many stationary sequences satisfy (2.2.2), including normal sequences with covariance sequence $\{r_n\}$ satisfying the "Berman Condition" $r_n \log n \to 0$. Sufficient conditions for values of $\theta < 1$ are given in [54], and an example

with $\theta=1/2$ appears later in this section. Examples can be found where the extremal index is zero, or does not even exist. This obviously has some theoretical interest but appears to occur in somewhat pathological cases and will not be pursued in the present discussion.

The usefulness of the extremal index appears from the following result.

Theorem 2.2.2 Suppose that the stationary sequence $\{\xi_n\}$ has extremal index $\theta>0$. Let $\{v_n\}$ be any sequence of constants and ρ any constant with $0\le\rho\le1$. Then $P\{\acute{M}_n \le v_n\} \to \rho$ if and only if $P\{M_n \le v_n\} \to \rho^\theta$.

This result makes no assumption about dependence, and is readily shown by obvious arguments (cf. [54] for details).

The following result now follows as a corollary, by taking $v_n=x/a_n+b_n$ in the theorem.

Theorem 2.2.3 Let the stationary sequence $\{\xi_n\}$ have extremal index $\theta>0$. If $P\{a_n(\acute{M}_n-b_n) \le x\} \to G(x)$ then $P\{a_n(M_n-b_n) \le x\} \to G^\theta(x)$ and conversely. That is $\acute{M}_n$ has an asymptotic distribution if and only if M_n does, with the power relation between the limits and the same normalizing constants.

By way of comment, note that G^θ is of the same type as G if one of them is of extreme value type (e.g. $[\exp(-e^{-x})]^\theta = \exp[-e^{-(x-\log\theta)}]$, and similarly for type II and III). If $\theta=1$ the limits for M_n and $\acute{M}_n$ are precisely the same. Indeed for $0<\theta<1$ the limits may also be taken to be the same by a simple change of normalizing constants.

The practical implication of this result is that one often does not need to be concerned about possible dependence in the data when applying classical extreme value theory. Indeed one may not have to worry about the

precise value of the extremal index since this only alters parameters of the
distribution which usually must be estimated in any case. Further, if $\theta > 0$,
the fact that the distributional type under dependence is the same as under
independence means that the classical domain of attraction criteria may be
applied to the marginal d.f. of the terms to determine which type applies.

The following simple example provides a case where $\theta < 1$, and will also
be useful later when the effects of the value of θ on the clustering of
exceedances will be discussed.

Example 2.2.4 Let $\eta_1, \eta_2 \ldots$ be i.i.d. with d.f. H and write $\xi_j = \max(\eta_j, \eta_{j+1})$.
Then $\{\xi_n\}$ is stationary with d.f. $F = H^2$ and an easy calculation shows that
if $u_n(\tau)$ satisfies (2.2.1) then $n[1-H(u_n(\tau))] \to \tau/2$ and

$$P\{M_n \leq u_n(\tau)\} = P\{\max(\eta_1, \ldots \eta_n) \leq u_n(\tau)\}\, P\{\eta_{n+1} \leq u_n(\tau)\} \to e^{-\tau/2}$$

so that $\{\xi_n\}$ has extremal index $\theta = 1/2$.

Criteria for determining the extremal index are discussed in [54].
Finally, we note that an interesting approach to the relating of dependent and
i.i.d. cases has been given recently by O'Brien [68]. This is based on the
general result

$$P\{M_n \leq u_n\} - F(u_n)^{nP\{\max(\xi_2, \xi_3, \ldots \xi_n) \leq u_n \mid \xi_1 > u_n\}} \to 0$$

which is shown in [68] to hold under weak dependence conditions, for a wide
variety of sequences $\{u_n\}$ and integers $p_n \to \infty$ with $p_n = o(n)$.

2.3 Relevant point process concepts.

In dealing with dependent cases it will be necessary to be somewhat
more formal than previously in the use of point process methods. Here we
establish the notation and framework (substantially following Kallenberg
([53])), and review a few key concepts which will be needed.

In general a point process is often defined on a locally compact second countable (hence complete separable metric) space S, though here S will invariably be a subset of the line or plane. Write S for the class of Borel sets on S and $B=B(S)$ for the bounded (i.e. relatively compact) sets in S. A point process ξ on S is a random element in $M=M(S)$, the space of locally finite (i.e. finite on $B(S)$) integer-valued measures on S where M has the vague topology and Borel σ-field $M = M(S)$.

Write $F = F(S)$ for the class of non-negative S-measurable functions, $\mu f = \int f d\mu$ for $\mu \in M$, $f \in F(S)$. The distribution $P\xi^{-1}$ of a point process ξ is uniquely determined by the distributions of $(\xi(I_1) \ldots \xi(I_k))$, $k=1,2\ldots,I_j \in T$ if T is any semiring whose generated ring is B. The distribution of ξ is also uniquely determined by the Laplace Transform

$$L_\xi(f) = Ee^{-\xi f}, \quad f \in F.$$

A (general) Poisson Process with intensity measure λ has the Laplace Transform $L_\xi(f)=\exp\{-\lambda(1-e^{-f})\}$ whereas a Compound Poisson Process has Laplace Transform

$$(2.3.1) \qquad\qquad L_\xi(f) = \exp\{-\lambda(1-L_\beta \circ f)\}$$

where β is a positive integer-valued random variable with Laplace Transform $L_\beta(t)=Ee^{\beta t}$. This consists of multiple events of (independent) sizes β located at the points of a Poisson Process having intensity measure λ.

Convergence in distribution of a sequence $\{\xi_n\}$ of point processes to a point process ξ is, of course, simply weak convergence of $P\xi_n^{-1}$ to $P\xi^{-1}$. It may be shown (cf. [53]) that $\xi_n \to \xi$ if and only if $L_{\xi_n}(f) \to L_\xi(f)$ for every $f \in F_c$, the subclass of F consisting of the nonnegative continuous functions with compact support. Point process convergence is also equivalent to convergence of finite dimensional distributions. Even more simply $\xi_n \overset{d}{\to} \xi$ if and only if $(\xi_n(I_1) \ldots \xi_n(I_k)) \overset{d}{\to} (\xi(I_1) \ldots \xi(I_k))$ $k=1,2\ldots I_j \in T$ where $T \subseteq B$

is a semiring such that $\xi(\partial B)=0$ a.s. for each $B\epsilon T$, and such that for any

$B\epsilon B$, $\epsilon > 0$, B may be covered by finitely many sets of T with diameter less

than ϵ (cf. [53 Theorem 4.2]). The semiclosed intervals and rectangles used

in Section 1.3 form such classes and hence e.g. (1.3.1) is indeed equivalent

to full convergence in distribution of N_k to N.

Finally the dependent counterpart of Theorem 1.3.4 requires the

concept of infinite divisibility. A point process ξ is said to be infinitely

divisible if for each $n=1,2,\ldots$ there exist some independent and identically

distributed point processes $\xi_1,\ldots,\xi_n$ such that $\xi \overset{d}{=} \xi_1 + \xi_2 \ldots + \xi_n$. The

Laplace Transform of an infinitely divisible point process has the canonical

representation

$$(2.3.2) \qquad - \log L\xi(f) = \int_{M\backslash\{0\}} (1-e^{-\mu f})\tilde{P}(d\mu)$$

where $\tilde{P}$ is a measure on $M\backslash\{0\}$ such that $\int_{M\backslash\{0\}} (1-e^{-\mu(B')})\tilde{P}(d\mu) < \infty$

for all $B\epsilon B$. $\tilde{P}$ is referred to as the canonical measure of ξ.

2.4 Convergence of point processes associated with extremes

We return now to the stationary sequence $\{\xi_n\}$ and consider point

process convergence results along the same lines as for the i.i.d. case in

Section 1.3. The notation of that and other previous sections will be used.

In particular N_n will denote the point process of exceedances on $(0,1]$ as

defined in Section 1.3, viz $N_n(E)=\#\{i/n\epsilon E: \xi_i > u_n, 1\leq i\leq n\}$, for a given

sequence of constants u_n.

When $\{\xi_n\}$ has extremal index $\theta=1$, the Poisson convergence result,

Theorem 1.3.1, may be proved provided $D(u_n)$ holds. This leads again to the

classical form (1.2.9) for the asymptotic distributions of extreme order

statistics. Similarly Theorem 1.3.2 holds under an r-level version $D_r(u_n)$ of

$D(u_n)$ (cf. [55, p. 107]) leading to the classical forms for the asymptotic joint distributions of extreme order statistics when $\theta=1$ (cf. Theorem 1.3.3). The "complete convergence" result Theorem 1.3.4 also holds giving again a Poisson limit in the plane when $\theta=1$ provided the multilevel conditions $D_r(u_n)$ all hold. These results are described in [55]; here we indicate the new features which arise when $0<\theta<1$.

As noted in Section 2.2, cases when $\theta<1$ occur when there is "high local dependence" in the sequence so that one exceedance is likely to be followed by others (see Example 2.2.4 as an illustration of this). The result is a clustering of exceedances, leading to a compounding of events in the limiting point process.

To include cases where such clustering occurs (i.e. $0<\theta<1$) we require the following modest strengthening of the $D(u_n)$ condition (cf.[50]).

Let $B_i^j(u_n)$ be the σ-field generated by the events $\{\xi_s \leq u_n\}$, $i \leq s \leq j$. For $1 \leq l \leq n-1$ write

(2.4.1) $\quad \beta_{n,l} = \max\{|P(A \cap B) - P(A)P(B)| : A \in B_1^k(u_n),\ B \in B_{k+l}^n(u_n),\ 1 \leq k \leq n-l\}$

Then the condition $\Delta(u_n)$ is said to hold if $\beta_{n,l_n} \to 0$ for some sequence l_n with $l_n = o(n)$. $\{\beta_{n,l}\}$ will be called the *mixing coefficients* for Δ. The condition Δ is of course stronger than D but still significantly weaker than strong mixing.

The condition Δ will be applied through the following lemma which is a special case of [84, Equation I').

Lemma 2.4.1. For each n and $1 \leq l \leq n-1$ write $\gamma_{n,l} = |\sup E\eta\zeta - E\eta\,E\zeta|$ where the supremum is taken over all η, ζ measurable with respect $B_1^j(u_n)$, $B_{j+l}^n(u_n)$ respectively, $0 \leq \eta, \zeta \leq 1$, $1 \leq j \leq n-l$). Then $\beta_{n,l} \leq \gamma_{n,l} \leq 4\beta_{n,l}$ where $\beta_{n,l}$ is the mixing coefficient for Δ, given by (2.4.1). In

particular $\{\xi_n\}$ satisfies $\Delta(u_n)$ if and only if $\gamma_{n,l_n} \to 0$ for some $l_n = o(n)$.

It will be convenient to have the following simple notion of clusters. Divide the $\{\xi_i\}$ into successive groups $(\xi_1,\ldots,\xi_{r_n})$, $(\xi_{r_n+1},\ldots,\xi_{2r_n})$ ... of r_n consecutive terms where r_n $(=o(n))$ is appropriately chosen. Then all exceedances of u_n within a group are regarded as forming a cluster. Note that since $r_n = o(n)$ the positions of the members of a single cluster will coalesce after the time normalization, giving nearly multiple points in the point process N_n on $(0,1]$. The following lemma shows that the clusters are asymptotically independent.

Lemma 2.4.2. Let $\tau > 0$ be constant and let $\Delta(u_n)$ hold with $u_n = u_n(\tau)$ satisfying (2.2.1). Suppose $\{k_n\}$ is a sequence of integers for which there exists a sequence $\{l_n\}$ such that $k_n l_n/n \to 0$ and $k_n \beta_{n,l_n} \to 0$, where $\beta_{n,l}$ is the mixing coefficient of $\Delta(u_n)$. Then, for each non-negative continuous f on $(0,1]$,

$$(2.4.2) \qquad E \exp(-\sum_{j=1}^{n} f(j/n)\chi_{n,j}) - \prod_{i=1}^{k_n} E \exp(-\sum_{j=(i-1)r_n+1}^{ir_n} f(j/n)\chi_{n,j}) \to 0,$$

where $\chi_{n,j}$ is the indicator $1_{\{\xi_j > u_n\}}$ and $r_n = [n/k_n]$.

This result is proved by the standard basic technique. Here the "intervals" $(1,\ldots,r_n)$, $(r_{n+1},\ldots,2r_n)$ ... are each shortened by omitting the final l_n integers of each, and successive approximations made for the first term of (2.4.2) in a similar way to the argument of Lemma 2.1.1, but using Lemma 2.4.1. (See [50] for details).

The number of exceedances in the ith cluster is $N_n((i-1)r_n/n, \; ir_n/n)$ $= \sum_{j=(i-1)r_n+1}^{ir_n} \chi_{n,j}$ and the cluster size distribution is therefore conveniently defined to be given by

(2.4.3) $\qquad \pi_n(i) = P\{\sum\limits_{j=1}^{r_n} \chi_{n,j}=i \mid \sum\limits_{j=1}^{r_n} \chi_{n,j} > 0\}, \qquad i=1,2,\ldots$

The following result gives sufficient conditions for N_n to have a Compound Poisson limit.

Theorem 2.4.3. Let the stationary sequence $\{\xi_n\}$ have extremal index $\theta>0$, and suppose that the conditions of Lemma 2.4.2 hold. If $\pi_n(i)$ (defined by (2.4.3)) has a limit $\pi(i)$ for each $i=1,2,\ldots$, then π is a probability distribution on $1,2,\ldots$ and the exceedance point process N_n converges in distribution to a Compound Poisson Process N with Laplace transform

(2.4.4) $\qquad L_N(f) = \exp \{-\theta\tau \int_0^1 (1- \sum\limits_{i=1}^{\infty} e^{-f(t)i} \pi(i)) \ dt\}$

Proof. The Laplace Transform $L_N(f)$ of N_n is precisely the first term of (2.4.2) and hence may be approximated by the second term. This latter term may be manipulated by using the facts that for large n, $f(j/n)$ is approximately constant in $(i-1)r_n < j \le ir_n$, and $Z_n = \sum\limits_{j=(i-1)r_n+1}^{ir_n} \chi_{nj}$ has the distribution $P\{Z_n=i\} = 1-p_n$ or $p_n\pi_n(i)$ according as $i=0$ or $i>0$, where $p_n=P\{N_n((0,r_n]) > 0\} \sim \theta\tau/k_n$ from Lemma 2.1.1.

The Laplace Transform (2.4.4) is of the form (2.3.1) with the integer valued r.v. β satisfying $P\{\beta=i\} = \pi(i)$ and intensity measure simply $\theta\tau m$ where m is Lebesgue measure. That is N consists of multiple events of size whose distribution is $\pi(i)$, located at the points of a Poisson Process having intensity $\theta\tau$.

The following result, showing that the Compound Poisson Process is the only possible limit for N_n under the conditions Δ is proved along similar lines to Theorem 2.4.3. (Full details may be found in [50]).

Theorem 2.4.4. Suppose $\tau>0$ is constant and the condition $\Delta(u_n)$ holds

$(u_n = u_n(\tau)$ satisfying (2.2.1)) for the stationary sequence $\{\xi_j\}$. If N_n converges in distribution to some point process N, then the limit must be a Compound Poisson Process with Laplace Transform (2.4.4) where π is some probability measure on $\{1,2,\ldots\}$ and $\theta = -\tau^{-1} \log \lim_{n \to \infty} P\{M_n \leq u_n(\tau)\} \in [0,1]$.

If $\theta \neq 0$ then $\pi\{i\} = \lim \pi_n\{i\}$ where π_n is defined by (2.4.3) for $r_n = [n/k_n]$, $k_n (\to \infty)$ being any sequence chosen as in Lemma 2.4.2.

Example 2.4.5. (Example 2.2.4 continued) It is evident that the exceedances of u_n by the process $\xi_j = \max(\eta_j, \eta_{j+1})$ in Example 2.2.4 occur in (at least) pairs, since if $\xi_{j-1} \leq u_n$ but $\xi_j > u_n$ then $\eta_{j+1} > u_n$ and hence $\xi_{j+1} > u_n$. It is readily seen by direct evaluation that $\pi_n(2) \to 1$ and hence $\pi(i) = 1$ or 0 according as $i = 2$ or $i \neq 2$. Thus the limiting point process N consists entirely of double events and (2.4.4) gives

$$L_N(f) = \exp\{-(\tau/2)\int_0^1 (1 - e^{-2f(t)})dt\}.$$

The most important application of the Compound Poisson limit is to give the asymptotic distribution of the kth largest value $M_n^{(k)}$ of $\xi_1 \ldots \xi_n$, when $\theta < 1$, using the relationship

(2.4.5) $$P\{M_n^{(k)} \leq u_n(\tau)\} = P\{N_n((0,1]) \leq k-1\}$$

Theorem 2.4.6. Suppose that for each $\tau > 0$, $\Delta(u_n)$ holds with $u_n = u_n(\tau)$ satisfying (2.2.1) and that $N_n(=N_n^{(\tau)})$ converges in distribution to some non-trivial point process $N (=N^{(\tau)})$ (which will occur e.g. if the conditions of Theorem 2.4.3 hold). Assume that the maximum M_n has the non-degenerate asymptotic distribution G as given in (1.2.1). Then for each $k = 1, 2, \ldots$

(2.4.6) $$P\{a_n(M_n^{(k)} - b_n) \leq x\} \to G(x) \left[1 + \sum_{j=1}^{k-1} \sum_{i=1}^{k-1} ((-\log G(x))^j / j!) \, \pi^{*j}(i)\right]$$

(zero if $G(x)=0$), where π_j is the j-fold convolution of the probability $\pi = \lim \pi_n$, π_n being given as in Theorem 2.4.4.

<u>Proof</u>. It follows from Theorem 2.4.4 that $\{\xi_n\}$ has extremal index $\theta>0$ and by Theorems 2.2.3 and 1.2.5 that $u_n(\tau) = a_n^{-1}x+b_n$ satisfies (2.2.1) with $\tau = \tau(x) = -\log G^{1/\theta}(x)$. The result follows using (2.4.5) since

$$P\{N_n((0,1]) \leq k-1\} \to P\{N((0,1]) \leq k-1\} = e^{-\theta\tau}\{1+ \sum_{j=1}^{k-1} (\theta\tau)^j \sum_{i=j}^{k-1} \pi^{*j}(i)/j!$$

which equals the right hand side of (2.4.6).

Note that the form (2.4.6) differs from the (classical) case $\theta=1$ (i.e. (1.2.9)), by inclusion of the convolution terms. These arise since e.g. the second largest may be the second largest in the cluster where the maximum occurs or the largest in some other cluster. This contrasts with the case $k=1$ for the maximum itself involving only the relatively minor change (Theorem 2.2.3) of replacing the classical limit by its θth power.

Finally in this section we indicate the modifications required by the dependence structure for "complete" convergence results such as Theorem 1.3.4. As in that case let N_n^* denote the point process in the plane consisting of points at $(j/n, u_n^{-1}(\xi_j))$ where u_n^{-1} denotes the inverse function of $u_n(\tau)$.

Under appropriate conditions (including e.g. that $\{\xi_n\}$ has extremal index $\theta=1$) N_n^* has again a Poisson limit N in $(0,\infty) \times (0,\infty)$ with Lebesgue measure as its intensity. However, as for the exceedance point process, the limit may undergo "compounding" when $\theta<1$.

The possible limiting forms for N_n^* were discussed first by Mori ([65]) under strong mixing conditions. More recently a transparent derivation has been given by Hsing ([48]) under weaker conditions, of $\Delta(u_n)$ type but involving multiple levels $u_n(\tau_i)$. A derivation similar to that for the

exceedance process shows that any limit in distribution of N_n^*, N^* say, must have independent increments, be infinitely divisible and have certain stationarity properties. These properties restrict the canonical measure $\tilde{P}$ of N^* to a form which can be readily determined (though requiring further notation), thus providing a specific expression for the Laplace Transform of N^*. Rather more illuminating, however, is the "cluster" representation of N^* which exhibits N^* as a Poisson Process in the plane together with a countable family of points with integer valued masses on vertical lines above and emanating from each Poisson point.

Specifically let $\delta_{(s,t)}$ denote unit mass at (s,t) and $\zeta = \sum_{i=1}^{\infty} \delta_{(S_i, T_i)}$ a homogeneous Poisson Process on $(0,\infty) \times (0,\infty)$. Let γ_i, $i=1,2,\ldots,$ be i.i.d. point processes on $[1,\infty)$ each independent of ζ, such that $\gamma_i(\{1\}) \geq 1, \gamma_i$ having points of mass a_{ij} at Y_{ij}. Then (under the assumed conditions),

$$N^* \stackrel{d}{=} \sum_i \sum_j a_{ij} \, \delta_{(S_i, T_i Y_{ij})}$$

As is clear from this representation N^* has atoms at each (S_i, T_i) (since the smallest $Y_{ij}=1$ for each i) and at points $(S_i, T_i Y_{ij})$ lying vertically above (S_i, T_i).

As noted in Sec. 1.3, theorems of this type summarize the relevant information concerning asymptotic joint distributions of extreme order statistics, in contrast to the individual marginal distributions obtained in Theorem 2.4.6.

2.5 Normal sequences: the comparison method.

For stationary normal sequences with covariances $\{r_n\}$, the condition $D(u_n)$ holds - as also does the sufficient condition (2.2.2) for the extremal index to be 1 provided the "Berman Condition" holds, viz.

(2.5.1) $\qquad r_n \log n \to 0 \qquad$ as $n \to \infty$.

These results are simply proved by means of a widely used comparison method which, in particular, bounds the difference between two (standardized) normal d.f.'s by a convenient function of their covariances. This result – here given in a general form – has been developed in various ways by Slepian [81], Berman [9] and Cramér (cf.[21]).

<u>Theorem</u> 2.5.1 (Normal Comparison Lemma). Suppose that $\xi_1, \ldots, \xi_n$ are standard normal random variables with covariance matrix $\Lambda^1 = (\Lambda^1_{ij})$ and $\eta_1, \ldots, \eta_n$ similarly, with covariance matrix $(\Lambda^0 = \Lambda^0_{ij})$, and let $\rho_{ij} = \max(|\Lambda^1_{ij}|, |\Lambda^0_{ij}|)$. Then, for any real numbers $u_1, u_2, \ldots, u_n$,

(2.5.2) $\quad P\{\xi_j \leq u_j, \; j=1,2,\ldots,n\} - P\{\eta_j \leq u_j, \; j=1,2,\ldots,n\} \leq$

$$(2\pi)^{-1} \sum_{1 \leq i < j \leq n} (\Lambda^1_{ij} - \Lambda^0_{ij})^+ \; (1-\rho^2_{ij})^{-1/2} \; \exp[-(u^2_i+u^2_j)/(2(1+\rho_{ij}))]$$

where $x^+ = \max(x,0)$. Further, replacing $(\Lambda^1_{ij} - \Lambda^0_{ij})^+$ by its absolute value on the right hand side of (2.5.2) yields an upper bound for the absolute value of the difference on the left hand side.

By taking $\xi_1, \xi_2 \ldots$ to be a stationary sequence of standardized normal r.v.'s with covariance sequence $\{r_n\}$ and $\eta_1 \; \eta_2 \ldots$ to be i.i.d. standard normal r.v.'s it follows simply from the theorem that if $\sup_n |r_n| < 1$ then for any real sequence $\{u_n\}$,

(2.5.3) $\quad |F_{i_1 \ldots i_s}(u_n) - \Phi^s(u_n)| \leq Kn \sum_{j=1}^{n} |r_j| \; \exp u^2_n/(1+|r_j|)$

where $F_{i_1 \ldots i_s}$ is the joint (normal) distribution of $\xi_{i_1}, \ldots, \xi_{i_s}$ and Φ is the standard normal d.f., $i_1 \ldots i_s$ being any choice of distinct integers from $1, 2 \ldots n$.

Now if $n(1-\mathcal{F}(u_n))$ is bounded and (2.5.1) holds it can be shown (by some routine calculation) that the right hand side of (2.5.3) tends to zero as $n \to \infty$, showing that $P\{\xi_{i_1} \le u_n, \ldots \xi_{i_s} \le u_n\}$ is approximately the same as it would be if the r.v.'s were i.i.d. instead of being correlated.

One can clearly (by identifying $i_1, \ldots i_s$ with $1, \ldots n$) then show directly that $P\{M_n \le u_n\}$ is approximately the same as for the i.i.d. standard normal sequence. Or Equation (2.5.3) may be simply used to verify the conditions $D(u_n)$, (2.2.2) and Theorem 2.2.3 used, thus leading by either route to the following result.

Theorem 2.5.2. Let $\{\xi_n\}$ be a (standardized) stationary normal sequence with covariances $\{r_n\}$ such that $r_n \log n \to 0$ as $n \to \infty$. Then

$$P\{a_n(M_n - b_n) \le x\} \to \exp(-e^{-x})$$

where a_n, b_n are given by (1.2.5).

Thus if $r_n \log n \to 0$, the maximum M_n from the stationary normal sequence has precisely the same asymptotic distribution as an i.i.d. normal sequence. The same is true of the distributions of all extreme order statistics. Although a slight weakening of (2.5.1) is possible this condition is close to being necessary for Theorem 2.5.2. Indeed if $r_n \log n \to \gamma > 0$ and $u_n = x/a_n + b_n$ (with a_n, b_n given by (1.2.5)) then the time normalized point processes of exceedances converge in distribution to a certain doubly stochastic Poisson Process. This leads to the asymptotic distribution of the maximum given by the convolution of a normal and Type 1 extreme value distribution. (See [55 Sec. 6.5] for details). Further, Mittal and Ylvisaker ([64]) have shown that if $r_n \to 0$ but $r_n \log n \to \infty$ then M_n has an asymptotic normal distribution. Thus in these "highly dependent" cases where $D(u_n)$ fails the classical theory no longer applies.

As noted previously stationarity has been assumed in many of the results to avoid the complications of notation and calculation which a nonstationary framework entails. For normal sequences, however, the sufficient correlation conditions still remain quite simple in nonstationary cases. For example the following result holds.

Theorem 2.5.3 Suppose that $\{\xi_n\}$ is a normal sequence with correlations r_{ij} satisfying $r_{ij} \leq \rho_{i-j}$ for $i \neq j$ where $\rho_n < 1$ for all n and $\rho_n \log n \to 0$ as $n \to \infty$. Let u_{ni} ($1 \leq i \leq n$, $n = 1, 2, \ldots$) be constants such that $\lambda_n = \min u_{ni} >$ $c (\log n)^{1/2}$ for some $c > 0$. If for some $\tau \geq 0$, $\sum_1^n (1 - \Phi(u_{ni})) \to \tau$, then $P\{\bigcap_{i=1}^n (\xi_i \leq u_{ni})\} \to e^{-\tau}$ as $n \to \infty$.

Theorem 2.5.3 has a very useful corollary in the case where a sequence $\{\eta_n\}$ is obtained from a stationary normal sequence $\{\xi_n\}$ by adding a varying mean – such as a seasonal component or trend. Calculations then show that the double exponential limit for the maximum still holds, but the normalizing constant b_n can require an appropriate modification. Specifically suppose that $\eta_i = \xi_i + m_i$ where $\{\xi_i\}$ is a standard (zero mean unit variance) normal sequence (not necessarily covariance stationary) and m_i are added deterministic components with the property that

$$(2.5.4) \qquad \beta_n = \max_{1 \leq i \leq n} |m_i| = o(\log n)^{1/2} \qquad \text{as } n \to \infty.$$

Under this condition it may be shown that a sequence of constants $\{m_n^*\}$ may be found such that

$$(2.5.5) \qquad \frac{1}{n} \sum_{i=1}^n \exp(a_n^*(m_i - m_n^*) - \frac{1}{2}(m_i - m_n^*)^2) \to 1 \qquad \text{as } n \to \infty$$

in which $a_n^* = a_n - \log \log n /(2a_n)$. With this notation, the following result

holds.

Theorem 2.5.4 Let $\eta_i = \xi_i + m_i$ as above where $\{\xi_n\}$ is a standard normal sequence with correlations r_{ij} satisfying $|r_{ij}| < \rho_{|j-i|}$ for $i \neq j$ with $\rho_n < 1$ and $\rho_n \log n \to 0$. Suppose that (2.5.4) holds and m_n^* satisfies (2.5.5). Then $M_n = \max (\eta_1, \eta_2, \ldots \eta_n)$ satisfies

$$P\{a_n(M_n - b_n - m_n^*) \leq x\} \to \exp (-e^{-x})$$

with a_n and b_n given by (1.2.5).

Thus the non-stationarity in the correlation structure has no effect on the limit law, and that introduced by the added deterministic component is adjusted for by the change of b_n to $(b_n + m_n^*)$. For details see [55, Chapter 6].

Normal processes provide a widely used source of models for describing physical phenomena, and it is gratifying that extremal theory applies so simply to them. Another convenient source of models is of course Markov chains, whose extremal behavior we discuss next.

2.6 Regenerative and Markov sequences

Most limit results for Markov chains are intimately tied to the theory of regenerative processes. For extreme values, this has been used in [2], [9], some further references on extremes of Markov chains being [10], [12], [68]. The "classical" case, exemplified by the GI/G/1 queue, is when a recurrent atom exists. However, recently regeneration techniques have been extended, in [6], [7], [67], to show that any Harris recurrent chain $\{\eta_n\}$ on a general state space is regenerative or 1-dependent regenerative (concepts to be defined below), and to give effective criteria for regeneration. Further,

clearly a function $\xi_n = f(\zeta_n)$ of a (1-dependent) regenerative sequence is (1-dependent) regenerative. An example where this added generality is useful is given by ARMA (p,q) - processes. They are naturally considered as functions of a Markov chain in R^{p+q}, and can be shown to be 1-dependent regenerative under weak conditions but usually not to be regenerative (cf. [79]).

Regenerative and 1-dependent regenerative sequences are strongly mixing, and hence the theory from Sections 2.1-2.4 applies, in particular the Extremal Types Theorem and the Compound Poisson limit for exceedances hold. However, this can also be obtained directly, and the direct approach gives some added insight, also into the results for general stationary sequences. In the present section this will be briefly outlined, along with some results directly tailored to Markov chains.

A sequence $\{\xi_t: t=1,2,\ldots\}$ is regenerative if there exists integer-valued random variables $0 < S_0 < S_1 < \ldots$ which divide the sequence into "cycles"

$$c_0 = \{\xi_n: 0 \le n < S_0\}, \quad c_1 = \{\xi_n: S_0 \le n < S_1\} \quad c_2 = \{\xi_n: S_1 \le n < S_2\}$$

which are independent and such that in addition $c_1, c_2, \ldots$ follow the same probability law. Then $\{S_k\}$ is a renewal process, i.e. $T_0 = S_0$, $T_1 = S_1 - S_0$, $T_2 = S_2 - S_1, \ldots$ are independent and $T_1, T_2, \ldots$ have the same distribution. We shall here assume that $m = ET_1 < \infty$ and that the distribution of T_1 is aperiodic, i.e. that the only integer for which $P(T_1 \in \{d, 2d, \ldots\}) = 1$ is $d=1$. The sequence $\{\xi_n\}$ is 1-dependent regenerative if there exists a renewal process $\{S_k\}$ as above, which makes $c_0, c_1, \ldots$ 1-dependent and $c_1, c_2, \ldots$ stationary.

Suppose now that $\{\xi_n: n=0,1,\ldots\}$ is a stationary regenerative sequence, let $\zeta_0 = \max \{\xi_i: 0 \le i < S_0\}$, $\zeta_1 = \max \{\xi_i: S_0 \le i < S_1\}$, $\zeta_2 = \max \{\xi_i: S_1 \le i < S_2\}$, $\ldots$ be the cycle maxima and define $\nu_t = \inf\{k \ge 1: S_k > t\}$. By the law of large numbers $\nu_t / t \to 1/m$ a.s. and $M_n = \max \{\xi_1, \ldots, \xi_n\}$ is easily

approximated by $\max \{\zeta_1, \ldots, \zeta_{\nu_n}\}$, which then in turn can be approximated by $\max \{\zeta_1, \ldots, \zeta_{[n/m]}\}$. Since $\zeta_1, \zeta_2, \ldots$ are i.i.d., this can be shown to lead to

(2.6.1) $$\sup_x |P(M_n \leq x) - G^n(x)| \to 0, \qquad \text{as } n \to \infty,$$

with $G(x) = P(\zeta_1 \leq x)^{1/m}$, see e.g. [79]. Since G is a d.f. it follows at once that the Extremal Types Theorem holds for $\{\xi_n\}$, and criteria for domains of attraction are obtained by applying the criteria for i.i.d. variables to $G(x)$.

In particular it follows from (2.6.1) and Theorem 1.2.5 that if $n(1-G(u_n)) \to \eta$ then

(2.6.2) $$P(M_n \leq u_n) \to e^{-\eta}, \qquad \text{as } n \to \infty,$$

and conversely if (2.6.2) holds then $n(1-G(u_n)) \to \eta$. As in Section 2.2 let $\xi_1', \xi_2', \ldots$ be the associated independent sequence which has the same marginal d.f. F as $\xi_1, \xi_2, \ldots$ and write $\hat{M}_n = \max \{\xi_1', \ldots, \xi_n'\}$. If in addition $n(1-F(u_n)) = nP(\xi_1 > u_n) \to \tau > 0$ then $P(\hat{M}_n \leq u_n) \to e^{-\tau}$ and $\{\xi_n\}$ hence has extremal index $\theta = \eta/\tau$. Since $1-G(u_n) \sim P(\zeta_1 > u_n)/m$, this can be formulated as follows. If there exists a sequence $\{u_n\}$ such that $n(1-F(u_n)) \to \tau > 0$ and

(2.6.3) $$\frac{P(\zeta_1 > u_n)/m}{P(\xi_1 > u_n)} \to \theta$$

then $\{\zeta_t\}$ has extremal index $\theta > 0$. In the same way it can be seen that conversely if $\{\zeta_t\}$ has extremal index $\theta > 0$ then for any $\tau > 0$ there exists a sequence $\{u_n\}$ which satisfies $n(1-F(u_n)) \to \tau$ and (2.6.3). Further, straightforward arguments show that (2.6.3) can be replaced by

(2.6.4) $$\lim_{x \uparrow x_F} \frac{P(\zeta_1 > x)/m}{P(\xi_1 > x)} = \theta$$

in this, with x_F the righthand endpoint of the d.f. F. However, it should
be noted that there are examples of regenerative sequences $\{\xi_t\}$ which satisfy
(2.6.2), even for $u_n = u_n(x) = x/a_n + b_n$, for all x, but for which
$(P(\int_1 > u_n)/m)/P(\xi_1 > u_n)$ does not converge, and hence the extremal index does
not exist, even if this is not expected to occur in cases of practical
interest.

A counterpart to the Compound Poisson limit Theorem 2.4.3 for the
exceedance point process N_n given by $N_n(E) = \# \{\frac{i}{n} \in E: \xi_i > u_n\}$, is also easy
to obtain for stationary regenerative sequences. Let N_n' be the point process
on $(0,1]$ which has points of multiplicity $\gamma_i = \#\{t: \xi_t > u_n, S_{i-1} \leq t < S_i\}$ at
i/n for each i for which $\gamma_i > 0$, i.e. N_n' is defined by $N_n'(E) = \sum_{i/n \in E} \gamma_i$.
Then $\{\gamma_i\}_{i=1}^\infty$ is an i.i.d. sequence, and if (2.6.2) holds so that
$nP(\gamma_1 > 0) = nP(\int_1 > u_n) \to \eta m$ and if

(2.6.5) $\qquad \pi_n(i) = P(\gamma_1 = i | \gamma_1 > 0) \to \pi(i)$, as $n \to \infty$,

for all i, for some $\{\pi(i); i=1,2,\ldots\}$ then it follows at once that N_n'
converges in distribution to a Compound Poisson process N' with Laplace
transform $L_{N'}(f) = \exp\{-\eta m \int_0^1 (1 - \sum_{i=1}^\infty e^{-f(t)i} \pi(i))dt\}$. By definition, a non-
zero γ_i corresponds to a cluster of γ_i exceedances of u_n by ξ_t for $S_{i-1} \leq t < S_i$,
and since $S_i/i \to m$ as $i \to \infty$ there is hence a cluster of γ_i points located
approximately at mi/n in N_n. Hence for an interval E, $N_n(E)$ is approximated
by $N_n'(m^{-1} \cdot E)$, (for $m^{-1} \cdot E = \{x: mx \in E\}$) and asymptotically $N_n(E)$ should have
the same distribution as $N'(m^{-1} E)$. This argument can easily be extended and
made stringent, to give the following result.

Theorem 2.6.1 Let $\{\xi_n: n=0,1,\ldots\}$ be a stationary aperiodic regenerative
sequence with $m < \infty$ which satisfies (2.6.4) and let $\{u_n\}$ be constants such
that (2.2.1), i.e. $n(1-F(u_n)) \to \tau$, and (2.6.5) hold. Then N_n converges in
distribution to a Compound Poisson Process N with Laplace transform (2.4.4),

i.e.

$$L_N(f) = \exp \{-\theta \tau \int_0^1 (1 - \sum_{i=1}^{\infty} e^{-f(t)i} \pi(i)) \, dt\}.$$

These results may also be extended to 1-dependent regenerative sequences, however with some extra complexity. Here we mention that the criterion (2.6.4) for the extremal index to be one then is replaced (cf. [78]) by

$$(2.6.6) \qquad \lim_{x \cdot x_F} \frac{P(\zeta_1 \geq x, \, \zeta_2 \leq x)/m}{P(\zeta_1 > x)} = \theta.$$

In [79], (2.6.6) is further used to find conditions for $\theta = 1$ for a function $\xi_t = f(\eta_t)$ of a Markov chain on a general state space. This result is expressed directly in terms of the transition probabilities $P_n(x) = P(f(\eta_1) > u_n \, \eta_0 = x) = P(\xi_1 > u_n | \eta_0 = x)$ as follows.

Theorem 2.6.2 Let $\{\eta_n\}$ be a stationary regenerative Markov chain with the cycle length T_1 aperiodic and satisfying $E \, T_1^\alpha < \infty$, for some $\alpha > 1$. If $u_n = u_n(\tau)$ satisfies (2.2.1) for some $\tau > 0$ and

$$E(P_n(\eta_0)^s) \, n^{1+s/\alpha} \to 0 \qquad \text{as } n \to \infty$$

for some $s > 1$ with $1/\alpha + 1/s < 1$ then $\{\xi_n\}$ has extremal index $\theta = 1$.

We also refer to [79, Theorem 4.1] and [68, Theorem 2.1] for additional results on the extremal index and Compound Poisson Convergence, for general distributionally mixing sequences, in a form which is particularly convenient for applications to Markov chains. Finally the restriction that the Markov chain (or regenerative sequence) is started with the stationary initial distribution is not essential. All the results hold for arbitrary initial distributions, provided only that

$$P(\zeta_0 > \max \{\zeta_1, \ldots, \zeta_k\}) \to 0 \qquad \text{as } k \to \infty.$$

2.7 Moving averages

Here, a stationary sequence $\{\xi_t\}$ is a moving average if it can be written in the form

$$(2.7.1) \qquad \xi_t = \sum_{i=-\infty}^{\infty} c_i \zeta_{t-i}, \qquad t = 0,\pm 1,\ldots,$$

where $\{\zeta_t\}$ is an i.i.d. sequence (the "noise sequence") and $\{c_i\}$ is a sequence of constants (the "weights") and where the sums are assumed to converge with probability one. If a stationary normal sequence has a spectral density – this holds e.g. if $\sum r_t^2 < \infty$, it can be represented in a non-unique way, as a moving average with normally distributed ζ's. Further, (2.7.1) includes the ARMA-processes (which satisfy a finite linear difference equation in the ζ's and hence are multi-dimensional Markov chains), which are extensively used in time series analysis. Thus, in particular, some of the themes from Sections 2.5 and 2.6 will be taken up again here, but from a slightly different point of view.

The extremal behavior of $\{\xi_t\}$ depends on both the weights and the two tails of the marginal d.f. of the noise variables in an intricate and interesting way. To reduce the amount of detail, we shall only describe the asymptotic distribution of the maxima, for the case of *non-negative* c_i's. The general case involves some extra complexity, since then an extreme negative noise variable which is multiplied by a negative c_i may contribute to a large ξ_t-value. In addition to this, the references cited below prove point process convergence and give rather detailed information on the sample path behavior near extremes, including the clustering which occurs when the extremal index is less than one. Here we will only exhibit the limiting form of the sample paths near extreme values without going into technicalities, referring to [75], [76], [29] for further details.

In cases when (1.2.1) holds, i.e. when

(2.7.2) $\qquad P\{a_n(M_n - b_n) \leq x\} \overset{d}{\to} G(x), \qquad$ as $n \to \infty$,

the asymptotic behavior of the maximum is specified by the constants $a_n > 0, b_n$

and the d.f. G. However, this involves an arbitrary choice, since if

a_n, b_n are replaced by a_n', b_n', where $a_n/a_n' \to a > 0$ and $a_n(b' - b_n) \to b$, then

(2.7.2) still holds, but with G(x) replaced by G(ax + b). In the sequel we

will keep the G's fixed, as the standard d.f.'s displayed in Theorem 1.2.1 and

hence describe extremal behavior by a_n, b_n and the type of G.

The effect of dependence on extremal behavior can be further

understood by comparing with extremes of the noise sequence and of the

associated i.i.d. sequence $\{\hat{\xi}_t\}$ with the same marginal d.f. as the moving

average $\{\xi_t\}$. Specifically, for $\tilde{M}_n = \max\{\zeta_1, \ldots, \zeta_n\}$ and $\hat{M}_n = \{\hat{\zeta}_1, \ldots, \hat{\zeta}_n\}$ there

are norming constants $\tilde{a}_n$, $\hat{a}_n > 0$ and $\tilde{b}_n$, $\hat{b}_n$ such that for the cases we

consider here,

(2.7.3) $\qquad P\{\tilde{a}_n(\tilde{M}_n - \tilde{b}_n) \leq x\} \to G(x)$

and

(2.7.4) $\qquad P\{\hat{a}_n(\hat{M}_n - \hat{b}_n) \leq x\} \to G(x),$

with the same G as in (2.7.2), and we shall indicate the relations between the

different norming constants.

The articles by Rootzen ([75]) and Davis and Resnick ([29]) are

concerned with noise variables which are in the domain of attraction of the

type II extreme value distribution, or equivalently when the noise variables

have a regularly varying tail,

(2.7.5) $\qquad P\{\zeta_0 > x\} = x^{-\alpha}L(x),$

with $\alpha > 0$, and L slowly varying at infinity. Hence, using the prescription

for norming constants given after Theorem 1.2.5, if γ_n satisfies $P\{\zeta_0 < \gamma_n\} \leq$

$1 - 1/n \leq P\{\zeta_0 \leq \gamma_n\}$, so that γ_n is roughly of the order $n^{1/\alpha}$, then (2.7.3)

holds, with

$$\begin{cases} \tilde{a}_n = \gamma_n^{-1}, \ \tilde{b}_n = 0 \\ G(x) = \exp(-x^{-\alpha}), \qquad x > 0. \end{cases}$$

Let $c_+ = \max \{c_i; \ i = 0, \pm 1, \dots\}$. Then also (2.7.2) is satisfied, with

(2.7.6)
$$\begin{cases} a_n = c_+^{-1} \tilde{a}_n, \ \tilde{b}_n = 0 \\ G(x) = \exp(-x^{-\alpha}). \end{cases}$$

This is elegantly proved in [29], by first noting that complete Poisson convergence of extremes of the $\int$-sequence is immediate (cf. Section 1.3) and then obtaining the corresponding result for the ξ's by a "continuous mapping" and approximation argument. [29] uses some summability assumptions on the c_i's, and for convenience that $c_i=0$ for $i=-1,-2,\dots$. However, it seems clear that the results hold without any restrictions beyond the assumption that the sums in (2.7.1) converge, cf. [75].

An intuitive explanation of (2.7.6) is that when the tails of the noise variables decrease slowly, as in (2.7.5), then the extreme noise values are very much larger than the typical ones, and that hence the maximal ξ_t-value asymptotically is achieved when the largest $\int_t$-value is multiplied by the largest weight, c_+. This of course agrees with (2.7.6), since the norming constants there are the same as those which apply to $\max\{c_+ \int_1, \dots, c_+ \int_n\}$. These heuristic arguments also easily lead to the following form of the normalized sample path $\xi_{t+\tau}/\xi_\tau$ near an extreme value at, say, the time point τ; asymptotically this ratio has the same distribution as the function Y_t given by

(2.7.7)
$$Y_t = U c_{-t},$$

where U is a certain random variable with values in the set $\{\dots 1/c_{-1}, 1/c_0, 1/c_1, \dots\}$. Thus, except for a random height, sample paths near extremes are

asymptotically deterministic.

The special case of (2.7.5) when the noise variables are stable (or "sum-stable", as opposed to max-stable) was studied first, in [75]. It has the appealing feature that then also the moving average, and indeed all linear functions of the noise variables are jointly stable. For such variables, it is easily seen that (2.7.4) holds, with

$$\begin{cases} \hat{a}_n = (\Sigma c_i^\alpha)^{1/\alpha} \, \tilde{a}_n, \; \tilde{b}_n = 0 \\ G(x) = \exp\{-x^{-\alpha}\}, \end{cases}$$

and hence also that the extremal index is $c_+^\alpha/\Sigma c_i^\alpha$, for the case of non-negative c's discussed here. Although not considered in [29], this can be shown to hold also for the general case (2.7.5), provided the sums involved are convergent.

The other class of moving averages which has been studied, in [76], is specified by

(2.7.8) $$P(\zeta_0 > x) \sim K x^\alpha e^{-x^p} \quad \text{as } x \to \infty,$$

where $K, p > 0$ and α are constants. Again it follows, using Theorem 1.2.5, that (2.7.3) holds, with

$$\begin{cases} \tilde{a}_n = p(\log n)^{1-1/p} \\ \tilde{b}_n = (\log n)^{1/p} + p^{-1}((\alpha/p)\log \log n + \log K)(\log n)^{1/p-1} \\ G(x) = \exp\{-e^{-x}\}. \end{cases}$$

Thus the center of the distribution of $\tilde{M}_n$ tends to infinity roughly as $(\log n)^{1/p}$, and the "scale parameter" a_n^{-1} is of the order $(\log n)^{1/p-1}$, which shows that for p>1 the distribution of $\tilde{M}_n$ becomes more and more concentrated as $n \to \infty$, and that it becomes increasingly spread out for 0<p<1, while the order of the scale does not change for p=1. As we shall see, the same holds for $\hat{M}_n$ and M_n.

The case when (2.7.8) holds with p=1 leads to intermediate behavior,

and we will only discuss the remaining cases. For $0<p<1$ again a large ξ-value is caused by just one large noise variable, in a similar way to the behavior when (2.7.5) holds. However, the non-zero $\tilde{b}_n$-terms cause some extra complications. Thus, (2.7.2) holds with

$$\begin{cases} a_n = c_+^{-1}\tilde{a}_n, \ b_n = c_+\tilde{b}_n \\ G(x) = \exp\{-e^{-x}\}, \end{cases}$$

in analogy with (2.7.6), but, writing k for the number of i's with $c_i = c_+$, the appropriate version of (2.7.4) involves

$$\begin{cases} \acute{a}_n = c_+^{-1}\tilde{a}_n, \ \acute{b}_n = c_+(\tilde{b}_n + (\log k)/\tilde{a}_n) \\ G(x) = \exp\{-e^{-x}\}. \end{cases}$$

Also the asymptotic form of the sample path $\xi_{t-\tau}/\xi_\tau$ near an extreme value at τ is similar. For $k = 1$ it is given by the deterministic function

$$y_t = c_{-t}/c_+,$$

while in the general case it is a random translate of this.

The case when (2.7.8) holds with $p>1$ is more intricate, since then an extreme ξ-value is caused by many moderately large noise variables in conjunction, and since extremal behavior is determined by the constant $||c||_q = (\Sigma|c_i|^q)^{1/q}$ and the function

(2.7.9) $$y_t = \Sigma_i c_{i-t} c_i^{q/p}/||c||_q^q,$$

with $q=(1-1/p)^{-1}$. In fact, the normalized sample path $\xi_{t+\tau}/\xi_\tau$ near an extreme at τ asymptotically has the deterministic form (2.7.9), and (2.7.2) and (2.7.4) hold, with

(2.7.10) $$\begin{cases} a_n = \acute{a}_n = ||c||_q^{-1}\tilde{a}_n, \ b_n = \acute{b}_n \\ G(x) = \exp\{-e^{-x}\}. \end{cases}$$

Here $b_n = \acute{b}_n$ is not determined by (2.7.8) alone, except for finite moving

averages, it is also influenced by the center of the distribution of the ζ's.
However, it is roughly of the order $||c||_q \tilde{b}_n$, but still
$a_n|(b_n-||c||_q \tilde{b}_n)|$ may in general tend to infinity. It of course follows at
once from (2.7.10) that the extremal index is one for p>1.

For p=q=2, which includes the normal case, (2.7.9) is the correlation
function and $||c||_q$ is proportional to the standard deviation, in agreement
with Section 2.5. However, it is interesting to note that for p≠2
covariances seem to have little bearing on extremes.

The results for the case (2.7.8) use the assumption that
$|c_i|=0(|i|^\theta)$, for some $\theta > $ max (1,2/q), and for p>1 in addition a number of
smoothness restrictions on the distribution of the noise variables. These are
mainly used in the derivations of the behavior of the tail of $\xi_0=\Sigma c_i \zeta_{-i}$,
which for p>1 is the main difficulty, cf. [78]. It is fairly easy to see that
$D(u_n)$ holds for all the moving averages considered here, and the results above
for p>1 are obtained along the lines set out in Section 2.2, by verifying
(2.2.2). For 0≤p≤1, i.e. in the cases when θ may be less than one, the
proofs use ad hoc methods, closely related to the heuristic arguments given
above.

Finally it should be mentioned that Finster ([33]) obtains some
related results using autoregressive representations of the processes, and
that Chernick ([18]) provides an example with qualitatively different
behavior.

2.8 Rates of convergence

Rates of convergence for the distribution of the maximum have mainly
been studied for i.i.d. variables. In the present section we briefly review
this work, discussing in turn pointwise rates, uniform convergence of d.f.'s,

so called "penultimate" approximations, uniform convergence over the class of all sets, and "large deviation" type results. Although generalizations seem straightforward, the only dependent sequences which have been considered are the normal ones. The quite precise results available for this case are discussed at the end of the section. A useful general observation, which applies to i.i.d. and dependent cases with extremal index $\theta=1$, is that once rates of convergence of the maximum have been found, then it is typically quite easy to find similar rates for k^{th} order statistics.

For i.i.d. random variables and a given u_n, the error $P(M_n \leq u_n) - e^{-\tau}$ in the approximation (1.2.3) is easy to compute directly, since then $P(M_n \leq u_n) = F^n(u_n)$, where F is the common d.f. of the variables. Further if F is continuous one can always make the difference zero for any $n, \tau > 0$ (by taking $u_n = F^{-1}(e^{-\tau/n})$). However, often u_n is determined from other considerations, e.g. in (1.2.1) it is chosen as $u_n = u_n(x) = x/a_n + b_n$ and correspondingly $\tau = \tau(x) = -\log G(x)$. Then the behavior of the approximation error

$$\Delta_n(x) = P(M_n \leq u_n(x)) - e^{-\tau(x)},$$

perhaps over a range of x-values, and in particular of

$$d_n(a_n, b_n) = \sup_x |\Delta_n(x)| = \sup_x |P(a_n(M_n - b_n) \leq x) - G(x)|$$

is less immediate. If (1.2.1) is used as an approximation or, more importantly if it motivates statistical procedures, when a_n, b_n have to be estimated, interest centers on which rate of decrease is attainable when the "best" a_n, b_n are used, i.e. on

$$d_n = \inf_{a>0, b} d_n(a,b) = \inf_{a>0, b} \sup_x |P(a(M_n - b) \leq x) - G(x)|.$$

It is easy to give examples of distributions F for which d_n tends to zero arbitrarily slowly, and to any exponential rate there is an F which achieves this rate. However faster than exponential decrease of d_n implies that F is

max-stable, and then $d_n=0$ for all n, [8], [77]. Also different standard distributions give quite different rates, e.g. for the normal distribution d_n is of the order $1/\log n$ while for the uniform and exponential distributions the order is $1/n$.

Let $\tau_n = \tau_n(x) = n(1-F(u_n(x)))$. In the sequel we will usually, for brevity, delete the explicit dependence on x. An obvious approach to analysing $\Delta_n(=\Delta_n(x))$ in the i.i.d. case is to introduce

$$\Delta'_n = (1-\tau_n/n)^n - e^{-\tau_n}, \quad \Delta''_n = e^{-\tau_n} - e^{-\tau},$$

so that

(2.8.1) $\quad |\Delta_n| = |F(u_n)^n - e^{-\tau}| = |(1-\tau_n/n)^n - e^{-\tau}| \le |\Delta'_n| + |\Delta''_n|$.

Here $0 \le \tau_n \le n$, and for such values the satisfying uniform bound

(2.8.2) $\qquad\qquad |\Delta'_n| \le n^{-1}(a + n^{-1})e^{-2}$

is derived by Hall & Wellner ([46]). Further, for fixed τ, by Taylor's formula

(2.8.3) $\qquad\qquad |\Delta''_n| \sim e^{-\tau} |\tau_n - \tau|$,

as $\tau_n \to \tau$. However, (2.8.3) is only uniform for $\tau_n - \tau = \tau_n(x) - \tau(x)$ in intervals which are bounded from below, and to bound d_n a further argument has to be added. Often this runs as follows; (2.8.2) and (2.8.3) give sharp estimates of $\sup_{x>a} |\Delta_n(x)|$, for any $a > x_0$, where x_0 is the left-hand endpoint of the d.f. G, and then also for $\sup_{x>x_n} |\Delta_n(x)|$ if x_n is taken to converge to x_0 suitably slowly. Combining this with

(2.8.4) $\qquad \sup_{x \le x_n} |\Delta_n(x)| \le \max\{F^n(x_n/a_n + b_n), G(x_n)\}$

leads to a bound for $d_n(a_n, b_n)$, and then by varying a_n, b_n, to bounds for d_n. This approach is used, with some variations, by Hall & Wellner ([46]), Davis ([26]), Cohen ([19] [20]), and Leadbetter et al. [55]). Here the bounds corresponding to (2.8.2) and (2.8.3) are asymptotically sharp, but there is a

possibility that λ_n' and λ_n'' can at least partially cancel. However, this happens only if $\tau_n = \tau - \tau^2/(2n) + o(1/n)$, and hence in fairly special cases, as is readily seen (cf. Davis ([26])).

A number of papers, some of the later references being Cohen ([19][20]), Smith ([82]), and Resnick ([73]), have introduced conditions which permit more explicit bounds than (2.8.1) - (2.8.4) to be calculated. Their approach is to take some set of conditions for attraction to an extreme value distribution, typically involving convergence of some quantity related to the tail of F, and show that if this holds at a specific rate then $d_n(a_n, b_n)$ converges at a corresponding rate. In this a set of simple sufficient conditions due to von Mises ([63]) (cf. [55], p. 16) have been particularly useful. There are many possible versions of such conditions, and hence many partially overlapping results have been obtained. As a typical example we cite the following result of Resnick ([73]).

Suppose F is differentiable and that there exists a continuous function g which tends monotonically to zero and which satisfies

$$(2.8.5) \qquad \left| \frac{xF'(x)}{F(x)\,(-\log F(x))} - \alpha \right| \le g(x), \qquad x > 0,$$

for some $\alpha>0$. Then, if a_n is chosen to satisfy $-\log F(a_n^{-1}) = n^{-1}$,

$$\sup_{x \ge 1} |F^n(x/a_n) - \exp\{-x^{-\alpha}\}| \le .2701\, g(a_n^{-1})/(\alpha - g(a_n^{-1})),$$

for n such that $g(a_n^{-1}) < \alpha$. Here (2.8.5) is a slight variation of von Mises' condition for attraction to the type II extreme value distribution, and the proof is somewhat different from the method outlined above, the main ingredient being an estimate of $-\log(-\log F(x))$. Resnick also obtains a somewhat more complicated bound for the supremum $d_n(a_n, 0)$ over all x.

For i.i.d. variables bounds on the rate of convergence of the maximum

automatically lead to bounds for the rate of convergence also of kth largest values. This follows by using (1.2.7) and any of the known bounds for the difference between the binomial and Poisson distributions, since S_n is binomial with parameters n, τ_n/n (see e.g. [55], Section 2.4).

The normal case, briefly mentioned above, of course has attracted special attention. Straightforward calculations show that for a_n, b_n given by (1.2.5),

$$\Delta_n(x) \quad [\exp(-e^{-x})e^{-x} (\log \log n)^2] /(16 \log n) \text{ as } n \to \infty$$

and in Hall ([44]) is shown that for i.i.d. normal variables there are constants $0 < c_1 < c_2 < 3$ such that $c_1/\log n \le d_n \le c_2/\log n$, for $n \ge 3$, i.e. the best rate of convergence is of the disconcertingly slow order $1/\log n$. However, this is partially offset by the fact that d_n is, nevertheless, fairly small for small n, e.g. for $n \le 10,000$ it compares well with the error in the normal approximation to the binomial distribution.

In their pioneering paper [34], Fisher & Tippet had already noticed the slow convergence rate for the normal case, and suggested improved "penultimate" approximations. The idea is that since the type I extreme value d.f. can be approximated arbitrarily well by type II (or type III) d.f.'s, if a d.f. can be approximated by a type I d.f., the same error can (in the limit) be achieved by a type II (or III) d.f., and there is always a possibility they can do better. This has been further developed by Cohen ([19] [20], who in particular shows that a penultimate approximation of the maximum of normal random variables by a type II extreme value d.f. improves the rate of convergence to $1/(\log n)^2$. The disadvantage with this approach is that the exponent α in the approximating d.f. then has to be chosen differently for different values of n. A related approach is to consider a function $|M_n|^a$ instead of M_n itself. This is pursued in Hall ([45]) and Haldane & Jayakar

([43]), and gives the rate of convergence $1/(\log n)^2$ for a=2, while other values of a lead to the same order $1/\log n$ as for M_n itself. Numerical computations show that these approximations also do better for small and moderate values of n, as could be expected.

A further statistically relevant question is to find rates of uniform convergence, i.e. to bound

$$d_n' = \inf_{a>0,b} \sup_{B \in \mathcal{B}} |P(a_n(M_n-b_n) \in B) - G(B)|$$

where $\mathcal{B}$ denotes the Borel sets in R, and $G(B)$ is the probability that a random variable with d.f. G belongs to B. The obvious approach is to bound the difference between the density (which is assumed to exist) of $a_n(M_n-b_n)$ and G'. Let $G'(x) = G(x)\gamma(x)$, so that $\gamma(x) = e^{-x}$, $\alpha x^{-\alpha-1}$, and $\alpha(-x)^{\alpha-1}$ for the type I, II and III extreme value distributions, respectively. Since (for i.i.d. variables),

$$\frac{d}{dx} P\{a_n(M_n-b_n) \leq x\} = \frac{d}{dx} F^n(x/a_n+b_n) = F(x/a_n + b_n)^{n-1} na_n^{-1}F'(x/a_n+ b_n),$$

where the first factor tends to G at a rate given by the references cited above, the main problem is to bound the difference $na_n^{-1} F'(x/a_n + b_n) - \gamma(x)$. The recent thesis by Falk ([32]) contains a survey of results in this direction, some further recent work being that of de Haan & Resnick ([42]) and Weissman ([86]).

Another problem which has attracted some attention, partly because of reliability applications, is the uniformity of the convergence of $P\{a_n(M_n-b_n)>x\}/(1-G(x))$ for large x; see Anderson ([2]) and de Haan & Hordijk ([40]).

For a stationary dependent sequence with extremal index $\theta=1$, a further source of error is the approximation by the associated independent sequence, i.e. the difference

$$\Delta_n''' (x) = P\{a_n(M_n - b_n) \le x\} - F^n(x/a_n + b_n)$$

where F is the marginal d.f. of the sequence. Cohen ([19]) shows, under weak covariance conditions, that for a stationary normal sequence Δ_n''' is $o(1/\log n)$, and hence that the rate of convergence in (1.2.1) is determined by the difference $F^n(x/a_n + b_n) - G(x)$, and hence is the same as in the i.i.d. case. Let ρ be the maximal correlation in the stationary normal sequence. Rootzen ([77]) gives a first order approximation and bounds for Δ_n''' which are roughly of the order $1/n^{(1-\rho)/(1+\rho)}$ for $\rho \ge 0$.

By using an embedding technique, these rates are extended also to $M_n^{(k)}$ and to point processes of exceedances. This embedding can be used more generally, and hence also in dependent cases rates for the maximum often easily lead to similar rates for k-th largest values.

2.9 Multivariate extremes

We shall discuss here only one multivariate problem, the Extremal Types Theorem for i.i.d. random vectors, and its extension to dependent sequences. As shown by de Haan & Resnick ([41]) and Pickands ([72]) the problem of characterizing the possible limit laws of the vector of coordinatewise maxima splits into two independent problems, to find the marginal d.f.'s which may occur - by the one-dimensional result this is just the class of extreme value d.f.'s - and to characterize the limiting dependence between components. Following Deheuvels ([30]) and Hsing ([49]) we will use the concept of dependence functions to discuss this.

Let $\underline{\xi} = (\xi_1, \ldots, \xi_d)$ be a d-dimensional random vector with d.f. $\underline{G}$ and marginal d.f.'s G_j, $1 \le j \le d$. The *dependence function* $\underline{D}$ of $\underline{\xi}$ (or of $\underline{G}$) is defined by

$$\underline{D}(x_1, \ldots, x_d) = P\{G_1(\xi_1) \le x_1, \ldots, G_d(\xi_d) \le x_d\}.$$

$\underline{D}$ is the d.f. of a distribution on $[0,1]^d$, and it has uniform marginal
distributions if the G_j's are continuous. The marginal distributions together
with the dependence function determines $\underline{G}$, since

(2.9.1) $\underline{G}(x_1,\ldots,x_d) = D(G_1(x_1),\ldots,G_d(x_d))$, $x_1,\ldots,x_d \in R$.

This is a consequence of the relation

$$\{G_j(\xi_j) \le G_j(x_j); \ 1\le j\le d\} \backslash \bigcup_{j=1}^{d} \{G_j(\xi_j) \le G_j(x_j), \ \xi_j > x_j\}$$

$$\subset \ \{\xi_j \le x_j; \ 1\le j\le d\} \subset \{G_j(\xi_j) \le G_j(x_j), \ 1\le j\le d\},$$

since it is readily seen that $P(G_j(\xi_j) \le G_j(x_j), \ \xi_j > x_j) = 0$ for each j.

A further useful property is that convergence of d-dimensional
distributions is equivalent to convergence of the dependence function *and* the
marginal distributions, provided the limit has continuous marginal d.f.'s.
This can be proved rather easily, using (2.9.1). Similarly to the one-
dimensional case, a d-dimensional d.f. $\underline{G}$ is said to be max-stable if there
exist constants $a_{n,i} > 0$, $b_{n,i}$, $i=1,\ldots,d$, such that

(2.9.2) $\underline{G}^n(a_{n,1}x_1 + b_{n,1},\ldots,a_{n,d}x_d + b_{n,d}) = \underline{G}(x_1,\ldots,x_d)$ $x_1,\ldots,x_d \in R$

for each $n=1,2,\ldots$. Further, a dependence function $\underline{D}$ is max-stable if

(2.9.3) $\underline{D}^n(x_1^{1/n},\ldots,x_d^{1/n}) = \underline{D}(x_1,\ldots,x_d)$, $x_1,\ldots,x_d \in R$,

for $n = 1,2,\ldots$.

Theorem 2.9.1 A d-dimensional ($d \ge 2$) d.f. with nondegenerate marginal
distributions is max-stable if and only if its marginal d.f.'s and its
dependence function are max-stable.

Proof. If $G_1,\ldots,G_d$ are max-stable, or if $\underline{G}$ is max-stable, then there are
constants $a_{n,i} > 0$, $b_{n,i}$ with $G_i^n(a_{n,i}x + b_{n,i}) = G_i(x)$, for $i=1,\ldots,d$. Hence,
in either case,

(2.9.4) $\underline{G}^n(a_{n,1}x_1+b_{n,1},\ldots,a_{n,d}x_d+b_{n,d}) = \underline{D}^n(G_1(a_{n,1}x_1+b_{n,1}),\ldots,G_d(a_{n,d}x_d+b_{n,d}))$

$$= \underline{D}^n(G_1(x_1)^{1/n},\ldots,G_d(x_d)^{1/n}).$$

Thus (2.9.2) follows at once if $\underline{D}$ is max-stable, by (2.9.1). The converse, i.e. that $\underline{D}^n(y_1^{1/n},\ldots,y_d^{1/n}) = \underline{D}(y_1,\ldots,y_d)$, for $y_i \in (0,1)$, $i=1,\ldots,d$ if G is max-stable also follows from (2.9.4), by taking $x_i = G_i^{-1}(y_i)$ there (note that each G_i is nondegenerate max-stable and hence continuous and strictly increasing on its support).

Let $\{\xi_n\} = \{(\xi_{n,1},\ldots,\xi_{n,d})\}_{n=1}^{\infty}$ be a sequence of i.i.d. random vectors, write $M_{n,i} = \max\{\xi_{1,i},\ldots,\xi_{n,i}\}$ and suppose there are constants $a_{n,i}>0$, $b_{n,i}$ such that

(2.9.5) $P\{a_{n,i}(M_{n,i}-b_{n,i}) \le x,\ 1 \le i \le d\} \overset{d}{\to} \underline{G}(x_1,\ldots,x_d)$,

where we may assume without loss of generality that the marginal distributions of $\underline{G}$ are nondegenerate. It then follows exactly as in the one-dimensional case that the possible limits $\underline{G}$ in (2.9.5) are precisely the max-stable d.f.'s. Thus, by Theorem 2.9.1 each marginal d.f. is max-stable and hence one of the three extreme value types, and the dependence function is max-stable. Further the distribution of $a_{n,i}(M_{n,i}-b_{n,i})$ tends to G_i, for $i=1,\ldots,d$ and the dependence function of $\{M_{n,i}: i=1,\ldots,d\}$ converges to the dependence function of $\underline{G}$. To complete the characterization of the limits, it only remains to describe the max-stable dependence functions. Again, this is a purely analytical problem, to solve the functional equation (2.9.3), and we thus only cite the result, which is obtained in somewhat varying forms in [45], [72], [30], and [49].

Theorem 2.9.2 A function $\underline{D}$ on $[0,1]^d$ is a max-stable dependence function if and only if it has the representation

$$\underline{D}(y_1,\ldots,y_d) = \exp\left(\int_S \min_{1 \le i \le d} \{x_i \log y_i\}\,du\right),$$

where S is the simplex $\{(x_1,\ldots,x_d): x_i \geq 0,\ i=1,\ldots,d,\ \sum_1^d x_i=1\}$, for some finite measure μ on S which satisfies $\int_S x_i d\mu = 1$, for $i=1,\ldots,d$.

Hsing, ([49]) also makes the observation that while the characterization of the limiting marginal d.f.'s is crucially tied to linear normalizations, this is not so for the dependence function. Specifically, if $\{u_{n,i}(x)\}$ are levels which are continuous and strictly increasing in x, and if

$$P(M_{n,i} \leq u_{n,i}(x_i),\ i=1,\ldots,d) \to \underline{G}(x_1,\ldots,x_d)$$

where $\underline{G}$ has continuous marginal distributions, then the dependence function of $\underline{G}$ is max-stable. The basic reason for this is the obvious fact that if $T_1,\ldots,T_d$ are continuous and strictly increasing, then $(\xi_1,\ldots,\xi_d)$ and $(T_1(\xi_1),\ldots,T_d(\xi_d))$ have the same dependence function.

Hsing also extends these results to stationary dependent sequences $\{\xi_n\}$, along rather similar lines as for the one dimensional case, as treated in Sections 2.1 and 2.2. Specifically, for given constants $\{u_{n,j}: j=1,\ldots,d, n \geq 1\}$ the condition $D(u_{n,1},\ldots,u_{n,d})$ is defined to hold if there is a sequence $1_n = o(n)$ such that $\alpha_{n,1_n} \to 0$ as $n \to \infty$ for

$$\alpha_{n,1} = \max \{|\, P(\xi_{i,j} \leq u_{n,j}: j=1,\ldots,d,\ i \in A \cup B)$$
$$-P(\xi_{i,j}\ u_{n,j}: j=1,\ldots,d,\ i \in A)\ P(\xi_{i,j} \leq u_{n,j}: j=1,\ldots,d,\ i \in B)\,|\}$$

where the maximum is taken over all sets A,B such that $A \subset \{1,\ldots,k\}$, $B \subset \{k+1,\ldots,n\}$, for some k. If $D(u_{n,1},\ldots,u_{n,d})$ holds the only possible limits in (2.9.5) again are the max-stable d.f.'s. Further, if in addition

$$(2.9.6)\quad \limsup_{n \to \infty} n \sum_{i_1=1}^{d} \sum_{i_2=1}^{d} \sum_{j=2}^{[n/k]} P(\xi_{1,i_1} > u_{n,i_1},\ \xi_{j,i_2} > u_{n,i_2}) \to 0 \text{ as } k \to \infty$$

then $P(M_{n,i} < u_{n,i},\ i=1,\ldots,d) \to \rho > 0$ if and only if $P^n(\xi_{1,i} \leq u_{n,i}: i=1,\ldots,d) \to \rho$, i.e. the asymptotic distribution of maxima is the same as if

the vectors were independent. ((2.9.6) of course reduces to (2.2.2) for d=1).

A further question considered by Hsing is independence of the marginals in the limiting distribution. In particular, he shows that if

$$(2.9.7) \quad \limsup_{\substack{i_1,i_2=1 \\ i_1 \neq i_2}}^{d} \sum_{j=1}^{[n/k]} P\{\xi_{1,i_1} > u_{n,i_1}, \xi_{j,i_2} > u_{n,i_2}\} \to 0,$$

as $k \to \infty$, and if $D(u_{n,1},\ldots,u_{n,d})$ is satisfied then (2.9.5) holds if and only if $P\{a_{ni}(M_{n,i}-b_{n,i}) \leq x\} \to G_i(x)$, as $n \to \infty$, for $i=1,\ldots,d$, and $\underline{G}$ then is of the form $\underline{G}(x_1,\ldots,x_d) = G_1(x_1)G_2(x_2) \cdots G_d(x_d)$.

Now let $\{\xi_n\}$ be normally distributed with $E\xi_{1,i}=0$, $V(\xi_{1,i})=1$ and let $r_{ij}(n)$ be the covariance between $\xi_{i,j}$ and $\xi_{1+n,i}$. If $r_{ij}(0) < 1$, for $1 \leq i \neq j \leq d$ and $r_{ij}(n) \log n \to 0$, as $n \to \infty$, for $i,j=1,\ldots,d$, and $u_{n,i}=x_i/a_n+b_n$, with a_n,b_n as in (1.2.5) then $D(u_{n,1},\ldots,u_{n,d})$, (2.9.6), and (2.9.7) are satisfied, so that the asymptotic distributions of maxima are the same as for a sequence of independent normal vectors with independent components (see [49], [1]).

2.10 Convergence of sums to non-normal stable distributions.

To consider the simplest case first, suppose $\{\xi_i\}$ is a sequence of i.i.d. symmetric r.v.'s, with $P\{\xi_i > x\} \sim x^{-\alpha}$, $x \to \infty$, for some $\alpha \in (0,2)$, so that the ξ_i's belong to the domain of attraction of the type II extreme value distribution, with norming constants $a_n=n^{-1/\alpha}$, $b_n=0$. Let $M_n^{(k)}$ and $m_n^{(k)}$ denote the kth largest and k^{th} respectively, of $\xi_1,\ldots,\xi_n$. It is straightforward to show that then

$$(2.10.1) \quad \limsup_{n \to \infty} P\{|a_n(\sum_{i=1}^{n}\xi_i - \sum_{k=1}^{r}(M_n^{(k)}+m_n^{(k)}))| > \delta\} \to 0 \quad \text{as } r \to \infty,$$

for any $\delta>0$. By Theorem 1.3.2 the joint distribution of $\{a_n M_n^{(k)}: k=1,\ldots,r\}$ converges, and the limit can be found explicitly, as in Theorem 1.3.3. For

example, the joint asymptotic distribution of $a_n M_n^{(1)}$ and $a_n M_n^{(2)}$ is as given there, with $G(x) = \exp(-x^{-\alpha})$. Let $\{M^{(k)}: k=1,\ldots,r\}$ have this distribution, so that $\{a_n M_n^{(k)}: k=1,\ldots,r\} \overset{d}{\to} \{M^{(k)}: k=1,\ldots,r\}$. Similarly, there are $\{m^{(k)}: k=1,\ldots,r\}$, with $\{a_n m_n^{(k)}: k=1,\ldots,r\} \overset{d}{\to} \{m^{(k)}; k=1,\ldots,r\}$, and it is easy to see that there is also joint convergence, $\{a_n M_n^{(k)}, a_n m_n^{(k)}: k=1,\ldots,r\} \overset{d}{\to} \{M^{(k)}, m^{(k)}; k=1,\ldots,r\}$. Hence

$$a_n \sum_{k=1}^{r} (M_n^{(k)} + m_n^{(k)}) \overset{d}{\to} \sum_{k=1}^{r} (M^{(k)} + m^{(k)}) \quad \text{as } n \to \infty.$$

It can further be shown that $\sum_{k=1}^{r} (M^{(k)} + m^{(k)})$ tends in distribution to a stable (or "sum stable", cf. Section 2.7), limit with index α as $r \to \infty$, and then it follows at once from (2.10.1) that $a_n \sum_{i=1}^{n} \xi_i$ has the same stable limit.

This illustrates that the central limit problem of convergence of sums to non-normal stable distributions hinges on the convergence of extreme order statistics, and the most natural approach to it is perhaps via extreme value theory. In Theorem 2.10.1 below this is made precise. The theorem, which builds on ideas of Durrett and Resnick ([31]) and Resnick ([73]), contains a functional central limit theorem, and the corresponding extreme value result is the "complete" convergence of upper and of lower extremes, which is discussed in Sections 1.3 and 2.4. The corresponding one-dimensional approach via the joint distribution of extremes, as sketched above, is used in [56] and [28] and will be briefly discussed at the end of this section.

The results depend essentially on the Ito-Lévy representation of the stable process, and we shall now list the relevant properties, referring to Ito ([51], Section 1.12) for proofs and further information. Let $\{\eta(t): 0 \le t \le 1\}$ be a non-normal stable stationary independent increments process (briefly, $\{\eta(t)\}$ will be referred to as a stable process). $\{\eta(t)\}$ can — and will throughout — be assumed to have sample paths in $D[0,1]$ the space of

functions on [0,1] which are right continuous and have left limits at each point. Let $S=[0,1] \times \overline{R}$, with $\overline{R}=[-\infty, \infty] \backslash \{0\}$, and define the Ito process N of jumps of $\{\eta(t)\}$ by

(2.10.2) $N(A) = \#\{t: (t, \eta(t) - \eta(t-)) \in A\}$

for Borel sets $A \subset S$, where $\eta(t) - \eta(t-)$ is the jump of $\eta(\cdot)$ at time t. Then N(A) is (measurable and) finite a.s. for each rectangle A such that $A \subset [0,1] \times [-\infty, -\epsilon] \cup [\epsilon, \infty]$ for some $\epsilon > 0$. Hence N is a point process, and in fact it is a Poisson process with intensity measure ν which is the product of Lebesgue measure and the measure ν' on $\overline{R}$ with density $\gamma_+ y^{-\alpha-1}$ for $y>0$ and $\gamma_- |y|^{-\alpha-1}$ for $y<0$, for some constants γ_+, $\gamma_- \geq 0$ which are not both zero (i.e. in shorthand notation, $d\nu = dt \times d\nu' = dt \times (\gamma_\pm |y|^{-\alpha-1} dy)$).

Let $m(\epsilon) = 0$ for $0<\alpha<1$, let $m(\epsilon) = \int_{\epsilon<|y|} y(1+y^2)^{-1} d\nu'(y)$ for $\alpha=1$,

and let $m(\epsilon) = \int_{\epsilon<|y|} y d\nu'(y)$ for $1<\alpha<2$, and define

(2.10.3) $\eta^{(\epsilon)}(t) = \iint\limits_{\substack{0 \leq s \leq t \\ \epsilon<|y|}} y dN - t m(\epsilon)$.

Here the integral is just a finite sum: if N has the points $\{(t_j, y_j): j>1\}$ then $|y_j| > \epsilon$ and $0 \leq t_j \leq 1$ only for finitely many j's, and

$$\iint\limits_{\substack{0 \leq s \leq t \\ \epsilon<|y_j|}} y dN = \sum_{\substack{j:\ t_j \leq t \text{ and} \\ \epsilon<|y|}} y_j.$$

With this notation

(2.10.4) $P(\sup_{0 \leq t \leq 1} |\eta(t) - \eta^{(\epsilon)}(t)| > \delta) \to 0$, as $\epsilon \to 0$,

for any $\delta > 0$.

Let $\{\xi_n\}_{n=1}$ be arbitrary random variables, let $\{a_n>0, b_n\}_{n=1}^{\infty}$ be norming constants, define stochastic processes: $\{\eta_n(t): 0<t<1\}_{n=1}^{\infty}$ in D[0,1] by

(2.10.5) $\eta_n(t) = \sum_{j=1}^{[nt]} a_n(\xi_j - b_n)$,

and in analogy with (2.10.2) let N_n be the point process of jumps of η_n, defined as

(2.10.6) $N_n(A) = \#\{t: (t, \eta_n(t) - \eta_n(t-)) \in A\} = \#\{j: (j/n, a_n(\xi_j - b_n)) \in A\}$,

for Borel sets $A \subset S = [0,1] \times \overline{R}$. The following theorem specifies the connection between convergence in distribution of η_n to η and of N_n to N. In this convergence is in $D[0,1]$ given the Skorokhod topology, see e.g. [16, Chapter 16].

Theorem 2.10.1. Let $\{\eta_n(t): 0 \le t \le 1\}$ and N_n be given by (2.10.5) and (2.10.6) and let $\{\eta(t): 0 \le t \le 1\}$ be a non-normal stable process with Ito process N defined by (2.10.2). Then $\eta_n \overset{d}{\to} \eta$ as $n \to \infty$, in $D[0,1]$, if and only if the following two conditions hold,

(2.10.7) $\qquad N_n \overset{d}{\to} N$ as $n \to \infty$, on S,

and, writing $I_{n,j} = 1$ if $|a_n(\xi_j - b_n)| > \epsilon$ and $I_{n,j} = 0$ otherwise

(2.10.8) $\limsup\limits_{n \to \infty} P\{ \sup\limits_{0 \le t \le 1} | \sum\limits_{j=1}^{[nt]} a_n(\xi_j - b_n)(1 - I_{n,j}) + tm(\epsilon)| > \delta \} \to 0$, as $\epsilon \to 0$,

for each $\delta > 0$.

Proof. Let $N^{(\epsilon)}$ and $N_n^{(\epsilon)}$ be the restrictions of N and N_n to $[0,1] \times$ $((-\infty, -\epsilon] \cup [\epsilon, \infty))$, for $\epsilon > 0$. Let $\eta^{(\epsilon)}$ be given by (2.10.3) and set

$$\eta_n^{(\epsilon)} = \int\limits_{\substack{0 \le s \le t \\ \epsilon < |y|}} y dN_n - tm(\epsilon) = \sum\limits_{j=1}^{nt} a_n(\xi_j - b_n) I_{n,j} - tm(\epsilon).$$

First, suppose that $\eta_n \overset{d}{\to} \eta$. The function which maps η into $N^{(\epsilon)}$ and η_n into $N_n^{(\epsilon)}$ is a.s. continuous with respect to the distribution of η (see Resnick ([73, p.   ]) and hence $N_n^{(\epsilon)} \overset{d}{\to} N^{(\epsilon)}$ for each $\epsilon > 0$. This implies that $N_n \overset{d}{\to} N$, i.e. (2.10.7) holds. Similarly,

$$\left| \eta_n(\cdot) - \eta_n^{(\epsilon)}(\cdot) \right| \overset{d}{\to} \left| \eta(\cdot) - \eta^{(\epsilon)}(\cdot) \right| \quad \text{in } D[0,1], \text{ and hence}$$

$$P\{ \sup_{0 \le t \le 1} |\eta_n(t) - \eta_n^{(\epsilon)}(t)| > \delta \} \to P\{ \sup_{0 \le t \le 1} |\eta(t) - \eta^{(\epsilon)}(t)| > \delta \} \quad \text{as } n \to \infty,$$

since $P(\sup_{0 \le t \le 1} |\eta(t) - \eta^{(\epsilon)}(t)| = \delta) = 0$, for $\delta > 0$. Now,

$$(2.10.9) \qquad \eta_n(t) - \eta_n^{(\epsilon)}(t) = \sum_{j=1}^{[nt]} a_n(\xi_j - b_n)(1 - I_{n,j}) + t m(\epsilon),$$

and (2.10.8) thus follows immediately from (2.10.4).

Conversely, suppose (2.10.7) and (2.10.8) hold. The map which takes N_n into $\eta_n^{(\epsilon)}$ is a.s. N-continuous, and hence $\eta_n^{(\epsilon)} \overset{d}{\to} \eta^{(\epsilon)}$, as $n \to \infty$ in $D[0,1]$, and together with (2.10.8), (2.10.9) this implies that $\eta_n \overset{d}{\to} \eta$, by [16], Theorem 4.2.

The main condition, $N_n \to N$, of "complete" convergence of extremes, requires much weaker asymptotic mixing conditions than those needed for convergence of sums to the normal distribution, cf. the end of Section 2.4. However, the local dependence restrictions, such as (2.2.2) may instead be rather restrictive, and are not even in general satisfied for 1-dependent processes, cf. Example 2.2.1.

The conditions of course become particularly simple when $\xi_1, \xi_2, \ldots$ are i.i.d. Then $N_n \to N$ is equivalent to $nP(a_n(\xi_1 - b_n) \in A) \to \nu'(A)$, for each Borel set $A \subset [-\infty, -\epsilon] \cup [\epsilon, \infty]$ for some $\epsilon > 0$, which in turn is the same as

$$(2.10.10) \qquad nP\{a_n(\xi_1 - b_n) > x\} \to \gamma_+ \int_x^\infty y^{-\alpha - 1} dy, \quad \text{for } x > 0,$$

$$nP\{a_n(\xi_1 - b_n) \le x\} \to \gamma_- \int_{-\infty}^x |y|^{-\alpha - 1} dy, \quad \text{for } x < 0,$$

as $n \to \infty$. Another way of expressing (2.10.10) is to say that the marginal d.f. F of the ξ's should belong to the domain of attraction of the type II

distribution for both maxima (if $\gamma_+ > 0$) and minima (if $\gamma_- > 0$), with the same norming constants $\{a_n > 0, b_n\}$. Furthermore, Resnick ([73]) shows that (2.10.10) actually implies also (2.10.8) for i.i.d. sequences. Thus in this case $\eta_n \overset{d}{\to} \eta$ in $D[0,1]$ is equivalent to (2.10.10). It may also be noted that b_n can be taken to be zero here.

If one is not interested in full convergence in $D[0,1]$, but only in "marginal" convergence of $\eta_n(1) = \sum_{j=1}^{n} a_n(\xi_j - b_n)$ to a non-normal stable distribution, sufficient conditions are easily found by "projecting onto the y-axis". Let N' be the point process of jump heights of η, given by

$$N'(A) = \#\{t \in [0,1]: \eta(t) - \eta(t-) \in A\} = N([0,1] \times A)$$

for Borel sets $A \subset \bar{R}$, so that N' is a Poisson process with intensity ν' and similarly let

$$N_n'(A) = \#\{j \in [1,n]: a_n(\xi_j - b_n) \in A\} = N_n([0,1] \times A).$$

By the same considerations as in the last part of the proof of Theorem 2.10.1, if

(2.10.11) $N_n' \overset{d}{\to} N'$,   as $n \to \infty$, in $\bar{R}$,

and if, as before with $I_{n,j} = 1$ if $|a_n(\xi_j - b_n)| > \epsilon$ and $I_{n,j} = 0$ otherwise,

$$\limsup P(|\sum_{j=1}^{n} a_n(\xi_j - b_n)(1 - I_{n,j}) + m(\epsilon)| > \delta) \to 0, \text{ as } \epsilon \to 0,$$

for each $\delta > 0$, then $\eta_n(1) \overset{d}{\to} \eta(1)$ in R. Moreover, it can be seen that (2.10.11) holds if and only if the joint distribution of the k largest and k smallest order statistics of $\xi_1, \ldots, \xi_n$ tends to the distribution of the k largest and k smallest jumps of $\{\eta(t): 0 \le t \le 1\}$ for each k, cf. the introduction to this section. This approach to convergence of $\sum a_n(\xi_j - b_n)$ to non-normal stable distributions is, with some variations, pursued in detail for i.i.d. ξ's by LePage, Woodroofe, & Zinn [(56)] and for stationary sequences satisfying distributional mixing conditions by Davis ([28]).

Finally, the results of this section easily carry over to non-stationary situations with [nt] replaced by an arbitrary time-scale, to convergence of row-sums in a doubly indexed array $\{\xi_{n,j}\}$ to a Lévy (independent increments) process without continuous component, to multi-dimensional ξ's, and also to convergence of so-called self-normalized sums.

2.11 Miscellanea

(a) Minima and maxima. Since the minimum $m_n = \min\{\xi_1, \ldots, \xi_n\}$ can be obtained as $m_n = -\max\{-\xi_1, \ldots, -\xi_n\}$, results for maxima carry directly over to minima. In particular it follows from the Extremal Types Theorem that, under distributional mixing assumptions, limiting d.f.'s of linearly normalized minima must be of the form $1-G(-x)$ where G is an extreme value d.f. Further, it is trivial to see that for i.i.d. variables minima and maxima are asymptotically independent (cf. [55], p. 28).

In a series of papers ([23], [24], [27]), R. Davis studies the joint distribution of m_n and M_n for stationary sequences $\{\xi_n\}$ under a number of different dependence restrictions. Here we only note that some of his results alternatively may be obtained as corollaries of the multivariate theory discussed in Section 2.9 by making the identification $\xi_{i,1} = \xi_i$, $\xi_{i,2} = -\xi_i$, so that $M_{n,1} = M_n$, $M_{n,2} = -m_n$. For example, writing $u_{n,1} = u_n$, $u_{n,2} = -v_n$, for $v_n \leq u_n$, the mixing condition $D(u_{n,1}, u_{n,2})$ then translates to $\alpha_{n,l_n} \to 0$ for some sequence $l_n = o(n)$, with

$$a_{n,l} = \max\{ \, |P(\xi_i \leq u_n, \, \xi_i \geq v_n : i \in A \cup B) - $$

$$P(\xi_i \leq u_n, \, \xi_i \geq v_n : i \in A) \, P(\xi_i \leq u_n, \, \xi_i \geq v_n : i \in B)| \, \},$$

where the maximum is taken over all sets $A \subset \{1, \ldots, k\}$, $B \subset \{k+1, \ldots, n\}$, for $k=1, \ldots, n-1$. Thus if this holds for $u_n = x/a_n + b_n$ and $v_n = y/c_n + d_n$, for all x and y it follows that any limiting d.f. of $(a_n(M_n - b_n), \, c_n(m_n - d_n))$ must be of

the form $\underline{G}(x,\infty)-\underline{G}(x,-y)$ where $\underline{G}$ is a bivariate extreme value d.f.

Furthermore the criterion (2.9.10) for independence of componentwise maxima, i.e. here for asymptotic independence of M_n and m_n translates to

$$\limsup_{n\to\infty} \sum_{j=2}^{[n/k]} \{P(\xi_1 > u_n,\ \xi_j < v_n) + P(\xi_1 < v_n,\ \xi_j > u_n)\} \to 0, \text{ as } k\to\infty.$$

(b) Poisson Limit Theorems. Although somewhat less generally formulated, the Poisson and Compound Poisson limits discussed in Section 2.4 amount to convergence of point processes N_n defined from a triangular array $\{\epsilon_{n,i}:$ $i=1,\ldots,n,\ n\geq 1\}$ of zero-one variables, with stationary rows $\epsilon_{n,1},\ldots,\epsilon_{n,n}$, by

$$N_n(E) = \sum_{i:i/n\epsilon E} \epsilon_{n,i},$$

for Borel subsets E of $(0,1]$. Thus, the proof of the Poisson limit for $\theta=1$ (see [55], Section 2.5) is easily seen to show that if $D(u_n)$ and (2.2.1) hold with $\xi_i \leq u_n$ and $\xi_i > u_n$ replaced by $\epsilon_{n,i}=0$ and $\epsilon_{n,i}=1$, respectively, then N_n converges to a Poisson process with intensity τ if and only if $nP(\epsilon_{n,1}=1) \to \tau$.

Conversely, the literature contains many sufficient conditions for convergence, which may be applied to extremes by setting $\epsilon_{n,i}$ equal to zero or one according to whether $\xi_i \leq u_n$ or $\xi_i > u_n$. Two sets of such conditions seem particularly useful here. For the first, let $B_{n,i}$ be the σ-algebra generated by $\epsilon_{n,1},\ldots,\epsilon_{n,i}$. Then the relation

$$(2.11.1) \qquad \sum_{i=0}^{[nt]} E\{\epsilon_{n,i+1}|\ B_{n,i}\} \to t\tau \qquad \text{as } n\to\infty,$$

in probability, for each $t\epsilon(0,1]$ is sufficient for convergence of N_n to a Poisson process with intensity τ ([36], [31]). For the second one, which is due to Berman ([12], [14]), we assume that each row has been extended to a doubly infinite sequence $\ldots,\epsilon_{n,-1},\epsilon_{n,0},\epsilon_{n,1},\ldots$ and write $\bar{B}_{n,i}$ for the

σ-algebra generated by $\epsilon_{n,1}, \ldots, \epsilon_{n,i}$. Then the relation

(2.11.1) $$\sum_{i=0}^{[nt]} E\{\epsilon_{n,i+1} | B_{n,i}\} \to t\tau \qquad \text{as } n \to \infty,$$

in probability, for each $t \in (0,1]$ is sufficient for convergence of N_n to a Poisson process with intensity τ ([36], [31]). For the second one, which is due to Berman ([12], [14]), we assume that each row has been extended to a doubly infinite sequence $\ldots, \epsilon_{n,-1}, \epsilon_{n,0}, \epsilon_{n,1}, \ldots$ and write $\overline{B}_{n,i}$ for the σ-algebra generated by $\ldots, \epsilon_{n,i-1}, \epsilon_{n,i}$. Berman's result is that if $nP(\epsilon_{n,1}=1) \to \tau$ and if there exists a sequence γ_n of integers, with $\gamma_n = o(n)$, such that

$$n \sum_{i=2}^{\gamma_n} P(\epsilon_{n,1}=1, \epsilon_{n,i}=1) \to 0, \quad n \to \infty,$$

$$nP(\epsilon_{n,1}=1 | \overline{B}_{n-\gamma_n}) \to \tau, \qquad n \to \infty,$$

in probability, then N_n again converges to a Poisson process with intensity τ.

Neither one of these three sets of conditions imply any of the others, in particular they are not necessary, and each of them might be the most convenient one in some situation. However, e.g. for normal sequences with $r_n \log n \to 0$ they all seem to lead to about the same amount of work. One useful feature of (2.11.1) is that it also directly gives rate of convergence results, cf. [77].

3. Extremes of continuous parameter processes.

3.1 The Extremal Types Theorem for stationary processes.

In this section we consider continuous parameter stationary processes and indicate the extremal results which are analogous to those of Chapter 2. Let, then, $\{\xi(t): t \geq 0\}$ be a strictly stationary process having a.s. continuous sample functions and continuous one-dimensional distributions. It

may then be simply shown as in [55, Chapter 7] (assuming that the underlying

probability space is complete) that $M(I) = \sup\{\xi(t): t \in I\}$ is a r.v. for any

finite interval I and, in particular, so is $M(T) = M([0,T])$. The extremal

types theorem may be proved even in this continuous context, showing that,

under general dependence restrictions, the only nondegenerate limits G in

(3.1.1) $P\{a_T(M(T) - b_T) \leq x\} \to G(x)$ as $T \to \infty$,

are the three classical types.

Though the general result requires considerable details of proof, the

method involves the very simple observation that for (any convenient) $h > 0$

(3.1.2) $M(nh) = \max(\zeta_1, \zeta_2, \ldots, \zeta_n)$

where $\zeta_i = \max\{\xi(t): (i-1)h \leq t \leq ih\}$. Thus if (3.1.1) holds and the

(stationary) sequence $\zeta_1, \zeta_2, \ldots$ satisfies $D(u_n)$ for each $u_n = x/a_{nh} + b_{nh}$, then

it follows from the discrete parameter Extremal Types result (Theorem 2.1.2)

that G must be one of the extreme value types. Hence the Extremal Types

Theorem certainly holds for strongly mixing stationary processes since then

the sequence $\{\zeta_n\}$ is also strongly mixing and thus trivially satisfies $D(u_n)$.

However a more general form of the theorem results from showing that the $D(u_n)$

condition holds for the ζ's when the ξ's satisfy certain conditions - in

particular a continuous version $C(u_T)$ of $D(u_n)$. In fact the condition $C(u_T)$

will be defined in terms of the process properties only at "time sampled"

points jq_T for a sampling interval $q_T \to 0$.

The ζ_i are of course maxima of $\xi(t)$ in fixed intervals of length h

(e.g. $\zeta_1 = M(h)$) and the sampling interval q_T must be taken small enough so

that these are well approximated by the maxima at the sample points jq_T. A

convenient restriction to achieve this is to define $q(=q(u))$ to satisfy

(3.1.3) $P(M(h) > u, \xi(jq) \leq u, 0 \leq jq \leq h) = o(\psi(u))$ as $u \to \infty$

where $\psi(u)$ is a function which will later be taken to represent the tail of

the distribution of M(h) but which for the present need only dominate $P\{\xi(0) > u\}$ i.e.

$$(3.1.4) \qquad\qquad P\{\xi(0) > u\} = o(\psi(u)).$$

In the following definition $F_{t_1 \ldots t_n}(u)$ will be written for $F_{t_1 \ldots t_n}(u, \ldots u)$, where $F_{t_1 \ldots t_n}(x_1, \ldots x_n) = P\{\xi(t_1) \le x_1, \ldots \xi(t_n) \le x_n\}$.

The *Condition* $C(u_T)$ will be said to hold for the process $\xi(t)$ and the family of constants $\{u_T: T > 0\}$, with respect to the constants $q_T \to 0$ if for any points $s_1 < s_2 \ldots < s_p < t_1 \ldots < t_p$, belonging to $\{kq_T: 0 < kq_T \le T\}$ and satisfying $t_1 - s_p \ge \gamma$, we have

$$\left| F_{s_1 \ldots s_p\ t_1 \ldots t_p}(u_T) - F_{s_1 \ldots s_p}(u_T)\ F_{t_1 \ldots t_p}(u_T) \right| \le \alpha_{T,\gamma}$$

where $\alpha_{T,\gamma_T} \to 0$ for some family $\gamma_T = o(T)$ as $T \to \infty$.

<u>Theorem</u> <u>3.1.1</u>. (Extremal Types Theorem for stationary processes) With the above notation suppose that (3.1.1) holds for the stationary process $\{\xi(t)\}$, and some constants a_T, b_T and a non-degenerate G. Suppose also that $\psi(u)$ is a function such that (3.1.4) holds and $T\psi(u_T)$ is bounded for $u_T = x/a_T + b_T$, for each x. If $C(u_T)$ holds for some family of constants $q_T = q(u_T)$ where $q = q(u)$ satisfies (3.1.3) then G must be one of the three classical extreme value types.

<u>Proof</u>. The method of proof is to take an arbitrary sequence of points $T_n \in (nh, (n+1)h]$, write $v_n = u_{T_n}$ and then relate $D(v_n)$ for the sequence ζ_n to $C(u_T)$ for the process $\{\xi_t\}$. This is achieved by approximating the joint distributions of the ζ_i by corresponding joint distributions of maxima at time sampled points jq_T. Details of the calculation may be found in [55, Section 13.1].

3.2 Domains of attraction

In the classical theory of extremes of i.i.d. sequences the type of limiting distribution for the maximum was determined by the asymptotic form of the tail of the distribution of ξ_1. This remained true for dependent stationary cases with non-zero extremal index since the limiting type was that of the associated independent sequence. For continuous parameter processes however it is clearly the tail of the distribution of ζ_1 (in view of (3.1.2)) rather than that of ξ_t which determines the limiting type. More specifically if $\zeta_1, \zeta_2 \ldots$ are i.i.d. random variables with the same distribution as $\zeta_1 = M(h)$ then $\{\zeta_n\}$ is called the *independent sequence associated with* $\{\xi_t\}$. If the ζ_n-sequence has extremal index $\theta > 0$ then any asymptotic distribution for $M(T)$ is of the same type as that for $\acute{M}_n = \max\{\zeta_1, \ldots, \zeta_n\}$. Again the case $\theta = 1$ is of special interest and sufficient conditions may be given. In particular the following condition (analogous to (2.2.2) for sequences) is useful:

The Condition $C'(u_T)$ will be said to hold for the process $\{\xi(t)\}$ and the family of constants $\{u_T: T > 0\}$ with respect to the constants $\{q = q(u_T) \to 0\}$ if

$$\limsup_{T \to \infty} (T/q) \sum_{h < jq < \epsilon T} P\{\xi(0) > u_T, \ \xi(jq) > u_T\} \to \cup \text{ as } \epsilon \to 0$$

We assume also as needed that for some function ψ

(3.2.1) $P\{M(h) > u\} \sim h\psi(u)$ as $u \to \infty$ for $0 < h < \delta$, some $\delta > 0$.

The following result may then be shown.

Theorem 3.2.1. Suppose that (3.2.1) holds for some function ψ and let $\{u_T\}$ be a family of constants such that $C(u_T)$, $C'(u_T)$ hold with respect to a family $\{q(u)\}$ of constants satisfying (3.1.3) with h in $C'(u_T)$ not exceeding $\delta/2$,

where δ is from (3.2.1). Then as $T \to \infty$

$$(3.2.2) \qquad\qquad T \psi(u_T) \to \tau > 0$$

if and only if

$$(3.2.3) \qquad\qquad P\{M(T) \leq u_T\} \to e^{-\tau}.$$

Proof. It is sufficient to show that (3.2.2) and (3.2.3) are equivalent when T is replaced by any sequence $T_n \in (nh, (n+1)h]$ and u_T by $v_n = u_{T_n}$. But it is readily seen that $P\{M(nh) \leq v_n\} - P\{M(T_n) \leq v_n\} \to 0$ so that it is sufficient to show equivalence of the relations $T_n \psi(v_n) \to \tau$ and $P\{M(nh) \leq v_n\} \to e^{-\tau}$.

Now the sequence $\{\zeta_n\}$ defined as in (3.1.2) satisfies $D(v_n)$ and it follows from Lemma 2.1.1 that for fixed $k=1,2,\ldots$, $n' = [n/k]$,

$$P\{M(nh) \leq v_n\} - P^k\{M(n'h) \leq v_n\} \to 0 \qquad \text{as } n \to \infty.$$

A further approximation of $P\{M(n'h) \leq v_n\}$ by $n'h\psi(v_n)$ may be obtained by approximating maxima over intervals with those from sampled points jq ([55, Corollary 13.2.2]) and the desired equivalences follow from obvious relations such as

$$1 - (\tau/k) - o(k^{-1}) \leq \liminf_{n \to \infty} P\{M(n'h) \leq v_n\} \leq \limsup_{n \to \infty} P\{M(n'h) \leq v_n\}$$

$$\leq 1 - (\tau/k) + o(k^{-1}).$$

This result has the following immediate corollary linking the asymptotic distributional properties of $M(T)$ to those of the maximum $\hat{M}_n = \max(\hat{\zeta}_1, \ldots, \hat{\zeta}_n)$ of the associated independent sequence $\{\hat{\zeta}_n\}$.

Corollary 3.2.2 Let $\{u_T\}$ be a family of constants such that the conditions of Theorem 3.2.1 hold. Let $0 < \rho < 1$. If

$$(3.2.4) \qquad\qquad P\{M(T) \leq u_T\} \to \rho \text{ as } T \to \infty$$

then

(3.2.5) $\qquad P\{\acute{M}_n \leq v_n\} \to \rho \text{ as } n \to \infty,$

with $v_n = u_{nh}$. Conversely if (3.2.5) holds for some sequence $\{v_n\}$ then (3.2.4) holds for any u_T such that $\psi(u_T) \sim \psi(v_{[T/h]})$ provided the conditions of Theorem 3.2.1 hold.

Proof. This follows simply from Theorems 3.2.1 and 1.2.5 by obvious identifications.

It may be seen simply from this how the function ψ can be used in the classical domain of attraction criteria to determine the type of limiting distributions G in (3.1.1) for M(T). In this for an extreme value d.f. G we write D(G) for the (classical) domain of attraction of G, i.e. the set of all d.f.'s F such that $F^n(x/a_n + b_n) \to G(x)$ for some sequences $\{a_n > 0\}$, $\{b_n\}$.

Corollary 3.2.3 Suppose that the conditions of Theorem 3.2.1 hold for all families of the form $u_T = x/a_T + b_T$ where $a_T > 0$ and b_T are given constants and that M(T) has the limiting distribution given by (3.1.1). Then

(3.2.6) $\qquad h\psi(u) \sim 1 - F(u) \qquad \text{as } u \to \infty$.

for some $F \in D(G)$. Conversely suppose that (3.2.1) holds, and (3.2.6) holds for some $F \in D(G)$. Let $a_n' > 0$, b_n' be constants such that $F^n(x/a_n' + b_n') \to G(x)$ and set $a_T = a'_{[T/h]}$, $b_T = b'_{[T/h]}$. Then (3.1.1) holds provided the conditions of Theorem 3.2.1 hold for each $u_T = x/a_T + b_T$.

Proof. This follows from the previous corollary, noting that if F is the d.f. of the associated independent sequence $\{\acute{\zeta}_n\}$ then $1 - F(u) = P\{M(h) > u\} \sim h\psi(u)$.

In particular if $\psi(u)$ satisfies one of the classical domain of attraction criteria when substituted for $1 - F(u)$, then the limiting distribution for M(T) is of that type. Thus $\psi(u)$ plays the central role in determining limiting types just as the tail $1 - F$ does in the discrete case.

3.3 Extremes of stationary normal processes

In this section we briefly indicate how the results apply to a stationary normal process $\xi(t)$ (assumed standardized to have zero mean, unit variances, and covariance function $r(t)$ satisfying

$$(3.3.1) \qquad r(t) = 1 - C|t|^\alpha + o(|t|^\alpha) \qquad \text{as } t \to 0$$

for some $C>0$, $0<\alpha\leq2$. This includes all the mean-square differentiable cases ($\alpha=2$) and a wide variety of cases with less regular sample functions ($0<\alpha<2$), such as the Ornstein-Uhlenbeck process ($\alpha=1$). It may be shown that for such a process (satisfying (3.3.1)) that a function $\psi(u)$ satisfying (3.2.1) is given by

$$(3.3.2) \qquad \psi(u) = C^{1/\alpha} H_\alpha u^{(2-\alpha)/\alpha} (2\pi)^{-1/2} e^{-u^2/2},$$

but the proof involves quite intricate computations when $\alpha<2$ (and indeed forms the main part of the total discussion of the asymptotic behavior of $M(T)$). The H_α are constants whose numerical values are known only in the cases $\alpha=1,2$ ($H_1=1$, $H_2=\pi^{-1/2}$). The "regular" case $\alpha=2$ is simpler and $\psi(u)$ may then be alternatively obtained as in the next section.

It can be shown using the Normal Comparison Lemma (Theorem 2.5.1) that the (standard) stationary normal process $\xi(t)$ satisfying (3.3.1) satisfies $C(u_T)$ if $T\psi(u_T) \to \tau \geq 0$ with ψ given by (3.3.2) provided that

$$(3.3.3) \qquad r(t) \log t \to 0 \text{ as } t \to \infty.$$

To show this, the required "sampling intervals" $q(u)$ are chosen to satisfy (3.1.3) with $qu^{2/\alpha} \to 0$ sufficiently slowly. It can also be shown quite readily that $C'(u_T)$ is satisfied under the same conditions.

The function $\psi(u)$ given by (3.3.2) satisfies the domain of attraction criteria for the Type 1 extreme value distribution (e.g. Theorem 1.2.4, with $\psi=1-F$). Indeed some calculation shows that (3.2.2) holds with $\tau=e^{-x}$, $u_T=x/a_T+b_T$, for

(3.3.4) $\qquad a_T = (2 \log T)^{1/2}$

$b_T = a_T + \{((2-\alpha)/2\alpha) \log \log T + \log(C^{1/\alpha} H_\alpha (2\pi)^{-1/2} 2^{(2-\alpha)/2\alpha})\}/a_T$

Hence Theorem 3.2.1 gives the following result.

<u>Theorem</u> 3.3.1. Let the (standardized) stationary normal process $\xi(t)$ have covariance function $r(t)$ satisfying (3.3.1) and (3.3.3). Then $P\{a_T(M(T) - b_T) \leq x\} \to \exp(-e^{-x})$ as $T \to \infty$, where a_T and b_T are given by (3.3.4).

This result was obtained by Cramér ([21]) for the case $\alpha = 2$ and a somewhat more restrictive condition on the rate of decay of $r(t)$ as $t \to \infty$. The result in its present generality was obtained by Pickands [69] with further subsequent refinements by other authors (see [55, Chap. 12, for references]). In particular considerable generality is afforded by the family of covariances satisfying (3.3.1), and the requirement $r(t) \log t \to 0$ imposes only a very mild assumption on the rate of convergence of $r(t)$ to zero as $t \to \infty$.

3.4 Finite upcrossing intensities, and point processes of upcrossings.

In the continuous parameter case exceedances of a level typically occur on intervals and do not form a point process. However a natural analog is provided by the upcrossings (i.e. points where excursions above a level begin) which can form a useful point process for discussing extremal properties. Further in many cases the intensity of this point process provides the function $\psi(u)$ needed for the determination of extremal type. Before proceeding it is of interest to note that an alternative to discussing upcrossings is to consider the amount of time which the process spends above a level. This approach, used by Berman, is briefly indicated in Section 3.7.

Let then (as before) $\{\xi(t): t \geq 0\}$ be stationary with a.s.

continuous sample functions, and continuous one-dimensional d.f. If u is a constant, $\xi(t)$ is said to have an *upcrossing of u at* $t_0 > 0$ if for some $\epsilon > 0$, $\xi(t) \leq u$ in $(t_0 - \epsilon, t_0)$ and $\xi(t) \geq u$ in $(t_0, t_0 + \epsilon)$. (Hence in particular $\xi(t_0) = u$.) Note also that $\xi(t)$ is (a.s.) not identically equal to u in any interval, so that $\xi(t) < u$ at (infinitely many) points of $(t_0 - \epsilon, t_0)$ and $\xi(t) > u$ at infinitely many points of $(t_0, t_0 + \epsilon)$.

Under the given assumptions the number $N_u(I)$ of upcrossings of u by $\xi(t)$ in an interval I is a (possibly infinite valued) r.v.. If $\mu(u) = EN_u((0,1)) < \infty$ then $N_u(I) < \infty$ a.s. for bounded I, and the upcrossings form a stationary point process N_u with intensity parameter $\mu = \mu(u)$.

For stationary normal processes satisfying (3.3.1) μ is finite when $\alpha = 2$ and is then given by *Rice's Formula*,

$$(3.4.1) \qquad \mu(u) = (C/2)^{1/2} \pi^{-1} e^{-u^2/2}$$

and for non-normal processes μ may be calculated under weak conditions as

$$(3.4.2) \qquad \mu(u) = \int_0^\infty z \, p(u,z) \, dz$$

where $p(u,z)$ is the joint density of $\xi(t)$ and its (q.m.) derivative $\xi'(t)$. In fact these relations can be shown simply since $\mu(u) = \lim_{q \downarrow 0} J_q(u)$ where

$$(3.4.3) \qquad J_q(u) = q^{-1} P\{\xi(0) \leq u < \xi(q)\} \qquad (q > 0).$$

Note that the calculation of μ as $\lim J_q(u)$ is potentially simple since $J_q(u)$ depends only on the bivariate distribution of $\xi(0)$ and $\xi(q)$. Under general conditions it is also the case, when $u \to \infty$ as $q \to 0$ in a suitably coordinated way that

$$(3.4.4) \qquad J_q(u) \sim \mu(u)$$

and

$$(3.4.5) \qquad P\{M(q) > u\} = o(\mu(u)).$$

It then follows that (3.1.3) holds if $\psi(u) = \mu(u)$. For

$$0 \leq (q\mu)^{-1} \, P\{\xi(0) \leq u, \ \xi(q) \leq u, \ M(q) > u\}$$

$$= (q\mu)^{-1} \, [P\{\xi(0) \leq u, \ M(q) > u\} - P\{\xi(0) \leq u < \xi(q)\}]$$

$$\leq (q\mu)^{-1} \, [P\{N_u(0,q) \geq 1\} - qJ_q] \leq (q\mu)^{-1} \, (q\mu - qJ_q)$$

$$= 1 - J_q/\mu \to 0$$

which with (3.4.5) readily gives (3.1.3). Also (3.2.1) is often satisfied in regular cases. Under such conditions it thus follows that $\psi(u)$ may be replaced by $\mu(u)$ in previous results such as Corollary 3.2.3. (For a precise statement of conditions see [55, Theorem 13.5.2]).

Thus the intensity $\mu(u)$ can provide a convenient means for determining the type of limiting distribution for $M(T)$. However the point process of upcrossings has further interesting properties analogous to those for exceedances in discrete parameter cases. In particular a Poisson limiting distribution may be obtained after suitable time normalization.

Specifically let $u=u_T$ and T tend to infinity in such a way that $T\mu(u_T) \to \tau > 0$. Define a normalized point process N_T^* of upcrossings having points at t/T when ξ has an upcrossing of u at t i.e. $N_T^*(I) = \#\{$upcrossings of u_T by $\xi(t)$ for $t/T \in I\}$. Then the following result holds.

Theorem 3.4.1 Suppose that the conditions of Theorem 3.2.1 hold, with $\psi(u)=\mu(u)$. Then N_T^* converges in distribution to a Poisson Process with intensity τ as $T \to \infty$. This in particular holds for the stationary normal processes satisfying (3.3.1) with $\alpha=2$ and (3.3.4).

Similar results may be obtained under appropriate conditions for the point process of local maxima of height at least u, as $u \to \infty$, leading in particular to the asymptotic distribution of $M^{(k)}(T)$, the kth largest local maximum in $[0,T]$. Indeed "complete Poisson convergence" results analogous to those indicated for sequences in Sections 1.3 and 2.4, may be obtained for the

point process in the plane consisting of the locations and heights of the local maxima. (cf. [55, Sections 9.5 and 13.6] for details).

Finally, it is also possible to obtain Poisson limits in cases with irregular sample paths when $\mu(u)=\infty$ (e.g. normal with $0<\alpha<2$) by the simple device of using the "ϵ-upcrossings" of Pickands [70] in lieu of ordinary upcrossings. Specifically, for given $\epsilon>0$, $\xi(t)$ has an ϵ-upcrossing of the level u at t_0 if $\xi(t) \leq u$ for $t \in (t_0-\epsilon, t_0)$, and $\xi(t)>u$ for some $t\epsilon(t_0,t_0+\eta)$, for each $\eta>0$, so that clearly the number of ϵ-upcrossings in a finite interval I is finite (indeed bounded by $(m(I)/\epsilon)+1$ where $m(I)$ is the length of I). This device was used in [69] to give one of the first proofs of Theorem 3.3.1.

3.5 χ^2-processes

The proofs for normal processes in Section 3.3, and also for the sequence case (Section 2.5) use the Normal Comparison Lemma (Theorem 2.5.1) in an essential way. It will also be the basis for the present section on functions $\{\chi(t)\}$ of stationary d-dimensional ($d \geq 2$) normal processes $\xi(t)=(\xi_1(t),\ldots,\xi_d(t))$ defined as

$$(3.5.1) \qquad \chi(t) = \sum_{i=1}^{d} \xi_i^2(t).$$

We shall assume that the components are standardized to have mean zero and the same variance one - here this is a real restriction and not just a question of normalization - and also that the components are independent. Then $\chi(t)$ has a χ^2-distribution and the process $\{\chi(t); t\geq 0\}$ is called a χ^2-process (with d degrees of freedom). Extremal properties of χ^2-processes, and of some related functions of $\xi(t)$, have been studied in detail by Sharpe ([80]), Aronowich and Adler ([4], [5]), and Lindgren ([2], [57], [59]). Here we will

follow the "geometrical" approach of [58], and use the fact that $\chi(t)$ is the radial part of $\xi(t)$ to find the asymptotic double exponential distribution of maxima of $\chi(t)$, referring the reader to [5] for results on minima. However, we will indicate how the results can be obtained quite smoothly from the general theory of Section 3.4, rather than by using Lindgren's direct calculations.

Now, suppose further that the component processes $\{\xi_i(t)\}$, $i=1,\ldots,d$ are continuously differentiable a.s., and have the same covariance function $r(t)$. We shall presently show that $\mu(u)$, the mean number of u-upcrossings by $\chi(t)$, $0 \leq t \leq 1$, is easily found from (3.4.2), and then apply Theorem 3.4.1. For $i=1,\ldots,d$, $\xi_i(0)$ and $\xi_i'(0)$ are jointly normal, and hence independent, since $\text{Cov}(\xi_i'(0), \xi_i(0)) = \lim_{h \to 0} E\{h^{-1}(\xi_i(h)-\xi_i(0))\xi_i(0)\} = r'(0)=0$, where the last equality holds because $r(t)$ is symmetric around zero. Similarly, if $\lambda=-r''(0)$ is the second spectral moment, $\xi_1'(t)$ has variance λ. Thus the conditional distribution of $\chi'(0) = \sum_{i=0}^{d} 2\,\xi_i(0)\xi_i'(0)$ given $\chi(0)= \sum_{i=1}^{d} \xi_i^2(0) = u>0$

is normal with mean zero and variance $\sum_{i=1}^{d} 4\lambda\xi_i^2(0)=4\lambda u$. Let $p(z|u)$ be the density of this conditional distribution and let $p(u)$ be the density of $\chi(0)$, i.e.

$$(3.5.2) \qquad p(u) = 2^{-d/2}\Gamma(d/2)^{-1}\,u^{d/2-1}e^{-u/2}.$$

Then, using (3.4.2), it follows that

$$\mu(u) = p(u) \int_0^{\infty} z\, p(z|u)\,dz$$

$$= 2^{-(d-1)/2}\,\Gamma(d/2)^{-1}(\lambda/\pi)^{1/2}\,u^{(d-1)/2}\,e^{-u/2},$$

for $u \geq 0$. For u fixed, $J_q(u)=P(\chi(0) \leq u < \chi(q))/q \to \mu(u)$ as $q \to 0$, and similarly (3.4.4) holds also when $u \to \infty$, $q \to 0$, with $u^{1/2}q \to a>0$ (cf. [58,

Lemma 3.5], viz. $J_q(u) \sim \mu(u)$ as $q \to 0$, as needed.

Theorem 3.5.1 Let $\xi(t)=(\xi_1(t),\ldots,\xi_d(t))$ be a continuously differentiable d-dimensional standardized normal process with independent components and the same covariance function $r(t)$, as above. Suppose further that $r(t) \log t \to 0$ as $t \to \infty$ and that

(3.5.3) $T\mu(u_T) \to \tau$ as $T \to \infty$,

and let N_T^* be the point process of upcrossings of u_T by $\{\chi(t/T): t \epsilon [0,1]\}$. Then N_T^* converges in distribution to a Poisson process with intensity τ, and in particular,

(3.5.4) $P\{\max_{0 \leq t \leq T} \chi(t) \leq u_T\} \to e^{-\tau}$, as $T \to \infty$.

Proof. We shall briefly indicate how the conditions of Theorem 3.4.1 can be checked. We assume that d=2, the extension to d>2 being straightforward. The main idea in [58] is to introduce the normal random field $\{\chi_\theta(t): 0 \leq \theta < 2\pi, t \geq 0\}$, where

$$\chi_\theta(t) = \xi_1(t)\cos\theta + \xi_2(t)\sin\theta$$

is the component of $\xi(t)$ in the direction $(\cos\theta, \sin\theta)$, and to note that then

(3.5.5) $\chi(t) = \sup_{0 \leq \theta < 2\pi} \chi_\theta(t)^2$

Thus $\sup_{0 \leq t \leq h} \chi(t) = \sup_{0 \leq t \leq h, \, 0 \leq \theta < 2\pi} \chi_\theta(t)^2$, and it follows at once from

the extremal theory for normal random fields that (3.2.1) holds, for $\psi(u)=\mu(u)$ and any h>0, see [58, Lemma 2.2]. As noted above for fixed a>0 (3.4.4) holds for $q = a/u^{1/2}$ and since clearly $P\{\xi(0) > u\} = o(\mu(u))$, (3.4.5) is satisfied and thus also (3.1.3) holds with $\psi(u) = \mu(u)$. Thus it only remains to establish $C(u_T)$ and $C'(u_T)$, for an arbitrary h, say h=1, and with this choice of $q = q(u)$ since all the conditions then also hold if a is taken to tend to zero sufficiently slowly as $u \to \infty$. For this we introduce a

further sampling, in the θ-direction, given by a parameter $r = r(u) = b/u^{1/2}$ with $b>0$. Let $\tilde{\chi}_r(t) = \max \{\chi_{ir}(t); i=0,\ldots,[2\pi/r]\}$. Then, by (3.5.5) and an easy geometrical argument,

$$(3.5.6) \qquad \chi(t) \cos^2 r \leq \tilde{\chi}_r(t)^2 \leq \chi(t),$$

for $0<r<\pi/2$. To show that $C'(u_T)$ holds let $u_T'=(u_T)^{1/2}\cos r$, so that by (3.5.6) and stationarity,

$$(3.5.7) \quad \frac{T}{q} \sum_{1\leq jq \leq \epsilon T} P(\chi(0)>u_T, \ \chi(jq) > u_T)$$

$$\leq \frac{T}{q} \sum_{1\leq jq \leq \epsilon T} P(\tilde{\chi}_r(0)>u_T', \ \tilde{\chi}_r(jq) > u_T')$$

$$\leq \frac{T}{q} \sum_{1\leq jq \leq \epsilon T} |P(\tilde{\chi}_r(0)>u_T', \ \tilde{\chi}_r(jq) > u_T') - P(\tilde{\chi}_r(0)>u_T')P(\tilde{\chi}_r(jq)>u_T')|$$

$$+ \epsilon (T/q)^2 \ P(\chi_r(0) > u_T')^2).$$

It is readily seen that $T\mu((u_T')^2) \to \tau' = \tau \exp\{b^2/2\}$ and that $\chi_\theta(t)$ has mean zero and variance one, and that $|\text{cov} (\chi_\theta(0), \chi_{\theta'}(t))| \leq |r(t)|$, for any θ,θ'. The Normal Comparison Lemma can then be routinely applied to show that the sum on the righthand side of (3.5.7) tends to zero. Further, since $\mu(u)/p(u) = (2\lambda u/\pi)^{1/2}$, by (3.5.2) and since $J_q(u) \quad \mu(u)$, it follows from (3.5.6) that

$$\epsilon (T/q)^2 \ P(\tilde{\chi}_r(0) > u_T')^2 \leq \epsilon (T/q)^2 \ P(\chi(0) > (u_T')^2)^2$$

$$= \epsilon (T\mu((u_T')^2))^2 \ (P(\chi(0) > (u_T')^2)/q\mu((u_T')^2))^2$$

$$\to \epsilon (\tau e^{b^2/2})^2 \ (8\lambda\pi)/a^2$$

and thus $C'(u_T)$ is satisfied.

Next, with the notation of $C(u_T)$,

$$(3.5.8) \qquad |F_{s_1,\ldots,s_p, \ t_1,\ldots,t_p} (u_T) - F_{s_1,\ldots,s_p} (u_T)F_{t_1,\ldots,t_p}(u_T)|$$

$$\leq \left| P(\tilde{\chi}_r(t) \leq \sqrt{u_T}: t \in \{s_1,\ldots,s_p, t_1,\ldots,t_{p'}\}) - \right.$$

$$P(\tilde{\chi}_r(t) \leq \sqrt{u_T}: t \in \{s_1,\ldots,s_p\}) \; P(\tilde{\chi}_r(t) \leq \sqrt{u_T}: t \in \{t_1,\ldots,t_{p'}\}) \left. \right|$$

$$+ \sum_{0 \leq jq \leq T} P(u_T \leq \chi(jq) \leq u_T/\cos^2 r).$$

Here the Normal Comparison Lemma may be applied, similarly as for $C'(u_T)$, to show that the first expression on the right tends to zero as $T \to \infty$, if $t_1 - s_p \geq \gamma_T$, for suitable $\gamma_T = o(T)$. Further, the last sum in (3.5.8) is bounded by

$$\frac{T}{q} P(u_T \leq \chi(0) \leq u_T/\cos^2 r) = \frac{T}{q}(e^{-2^{-1}u_T} - e^{-2^{-1}u_T/\cos^2 r})$$

$$\to (2\pi/\lambda)^{1/2} \tau (1 - e^{-b^2/4})/a \qquad \text{as } T \to \infty,$$

by straightforward computations. Since this limit tends to zero as $b \to 0$, for a fixed, this may be seen to prove $C(u_T)$.

It is easy to "solve" (3.5.3), to show that (3.5.4) implies that

$$P\{a_T (\max_{0 \leq t \leq T} \chi(t) - b_T) \leq x\} \to \exp(-e^{-x}) \qquad \text{as } T \to \infty,$$

for

$$a_T = 1/2, \quad b_T = 2 \log T + (d-1) \log\log T - \log (\Gamma(d/2)^2 \pi/\lambda).$$

It might also be noted that this proof of $C(u_T)$ and $C'(u_T)$ applies without change also when the components of $\underline{\xi}(t)$ are dependent and have different covariance functions.

3.6 Diffusion processes

Diffusion processes have many useful special properties, and correspondingly several different approaches to their extremal behaviour are possible. E.g. Darling and Siegert ([22]), Newell [66], and Mandl [62] apply

transform techniques and the Kolmogorov differential equations (cf. also the survey [17]), Berman [11] exploits the regenerative nature of stationary diffusions, similarly to Section 2.6, and Davis [25] and Berman [13] use a representation of the diffusion in terms of an Ornstein-Uhlenbeck process. Here we shall discuss some aspects of Davis' methods, and in particular state his main result (relation 3.6.6) below).

A diffusion process $\{\xi(t); t \geq 0\}$ can be specified as the solution of a stochastic differential equation

$$(3.6.1) \qquad d\xi(t) = \mu(\xi(t))dt + \sigma(\xi(t))dB(t),$$

where $\{B(t): t \geq 0\}$ is a standard Brownian motion. We refer to [52] for the precise definition, and for the properties of $\{\xi(t)\}$ used below. For simplicity we will consider a somewhat more restrictive situation than in [52], and in particular we assume that $\{\xi(t)\}$ is defined on some open, possibly infinite, interval $I=(r_1,r_2)$ and that μ and σ are continuous, with $\sigma > 0$ on I.

Let $\{s(x); x \in I\}$ be a solution of the ordinary differential equation

$$(3.6.2) \qquad \sigma^2(x)s''(x) + 2\mu(x)s'(x) = 0$$

i.e. let it have the form $s(x)=c_1+c_2 \int_{x_0}^{x} \exp\{-\int_{x_0}^{y}(2\mu(z)/\sigma^2(z))dz\}\,dy$, with $c_2 > 0$, c_1 real constants, for some point $x_0 \in I$. Then s is strictly increasing and by Ito's formula $\eta_t = s(\xi_t)$ satisfies $d\eta_t = f(\eta_t)dB_t$, for $f(x) = s'(s^{-1}(x))\sigma(s^{-1}(x))$, i.e. s is a *scale function* and $\{\eta_t: t \geq 0\}$ is the diffusion on natural scale. The *speed measure*, m, corresponding to this scale function then has density $1/f(x)$, i.e. $m(dx) = (1/f(x))dx$. We further assume that the speed measure is finite, $|m| = \int_I m(dx) = \int_I (1/f(x)dx < \infty$, and that $s(x) \to \infty$ as $x \to r_2$, $s(x) \to -\infty$ as $x \to r_1$. It then follows that the boundaries r_1, r_2 are inaccesible, that the diffusion is recurrent, and that there exists a stationary distribution so that $\{\xi(t)\}$ becomes a stationary process if $\xi(0)$ is given this distribution.

The Ornstein-Uhlenbeck process, which will be denoted by $\{\tilde{\xi}(t)\}$ here, is

the stationary diffusion process (3.6.1) specified by $I=R$, $\mu(x)=x/2$, $\sigma(x)=1$, $x \in I$. For the present purposes, a convenient choice of scale function for $\{\tilde{\xi}(t)\}$ is $\tilde{s}(x)=(2\pi)^{1/2}\int_0^x e^{y^2/2}\,dy$, and the corresponding speed measure is $\tilde{m}(dx) = (2\pi)^{-1/2}e^{-x^2/2}\,dx$. Further, it can be seen that $\{\tilde{\xi}(t)\}$ is a standardized stationary normal process with covariance function $r(t)=e^{-|t|}$ and that $\tilde{s}(x) \sim (2\pi)^{1/2}x^{-1}e^{x^2/2} = (x\phi(x))^{-1}$, as $x \to \infty$. Hence, Theorem 3.3.1 may be applied with $C=\alpha=1$ and its conclusion can, e.g. by a simple "subsequence argument" be written as

$$(3.6.3) \qquad \sup_{u>u_0}\ |P(\tilde{M}(T) \le u) - e^{-T/\tilde{s}(u)}| \to 0 \quad \text{as } T \to 0,$$

for any $\tilde{u}_0>0$, and with $\tilde{M}(t)=\sup\{\tilde{\xi}(t): 0 \le t \le T\}$.

The main additional fact needed is that the Ornstein-Uhlenbeck process on natural scale can, by a change of time, be made to have the same distribution as $\{\eta(t)\}$. More precisely, ([25, Theorems 2.1 and 2.2]), there exists a strictly increasing random function $\{\tau(t): t \ge 0\}$ such that the processes $\{s(\xi(t)): t \ge 0\}$ and $\{\tilde{s}(\tilde{\xi}(\tau(t))): t \ge 0\}$ have the same distribution, and which satisfies

$$(3.6.4) \qquad T^{-1}\tau(T) \to 1/|m| \qquad \text{as } T \to \infty,$$

almost surely.

As in Section 2.6 it follows easily from (3.6.3), (3.6.4) that

$$(3.6.5) \qquad \sup_{u>u_0}\ |P(\tilde{M}(\tau(T)) \le u) - e^{-T/(\tilde{s}(u)|m|)}| \to 0, \text{ as } T \to \infty.$$

Since for $M(T) = \sup\{\xi(t): 0 \le t \le T\}$,

$$P(M(T) \le u) = P(\sup\{s(\xi(t)): 0 \le t \le T\} \le s(u))$$

$$= P(\sup\{\tilde{s}(\tilde{\xi}(\tau(t))): 0 \le t \le T\} \le s(u))$$

$$= P(\tilde{M}(\tau(T)) \le \tilde{s}^{-1}(s(u))),$$

(3.6.5) is readily seen to imply the main result of [27], that

$$(3.6.6) \qquad \sup_{u<u_0}\ |P(M(T) \le u)-e^{-T/(s(u)|m|)}| \to 0 \quad \text{as } T \to \infty,$$

for any $u_o \in I$ with $s(u_o) > 0$. This is a quite explicit description of $M(T)$, "as the maximum of T i.i.d. random variables with d.f. $G(u) = \exp\{-1/(s(u)|m|)\}$", and in particular domains of attraction for $M(T)$ are found by applying the classical criteria to $\exp\{-1/(s(u)|m|)\}$. Finally, as for Markov chains, the hypothesis of stationarity is not essential, (3.6.6) holds for any initial distribution, as can be seen e.g. by a simple "coupling argument".

3.7 Miscellanea

(a) Moving averages of stable processes. These are continuous time processes of the form $\xi(t) = \int c(t-x) d\zeta(x)$, with $\{\zeta(x)\}$ a non-normal stable independent increments process. Their extremal behaviour, which is similar to that of the corresponding discrete parameter moving average (cf. Section 2.7), is studied in detail in [75].

(b) Sample path properties. As mentioned in Section 2.7, [75], [76] and [29] also study the asymptotic distribution of sample paths near extremes. A different approach to this problem, via so-called Slepian model processes, has been pursued by G. Lindgren in a series of papers, cf. the survey [60] and the references therein.

(c) Extremal properties and sojourn times. In an important series of papers, Berman studies "the sojourn of $\xi(t)$ above u", defined as $L_T(u) = \int_0^T 1(\xi(t) > u) dt$, where $1(\cdot)$ is the indicator function. For a wide variety of cases, including many normal processes, χ^2-processes, Markov processes, and random Fourier sums, he finds the asymptotic form of the distribution of $L_T(u)$ as $u \to \infty$ for fixed T, and as $u, T \to \infty$ in a coordinated way. Further, he uses the equivalence of the events $\{M(T) > u\}$ and $\{L_T(u) > 0\}$ to study the maximum of $\{\xi(t)\}$. This work is reviewed in the present journal ([12]) by Berman himself.

REFERENCES

Because of space limitations we have not attempted to give a complete set of references. For further references we refer to the bibliographies complied by Harter [47] and Tiago de Oliveira and coworkers [83] (cf. also [54]).

[1] Amram, F. (1985). Multivariate extreme value distributions for stationary Gaussian sequences. J. Multivariate Anal. 16, 237-240.

[2] Anderson, C.W. (1970). Extreme value theory for a class of discrete distributions with applications to some stochastic processes. J. Appl. Probab. 7, 99-113.

[3] Anderson, C.W. (1978). Super-slowly varying functions in extreme value theory. J.R. Statist. Soc. B 40, 197-202.

[4] Aronowich, M. and Adler, R.J. (1985). Behaviour of χ^2 processes at extrema. Adv. Appl. Probab. 17, 280-297.

[5] Aronowich, M. and Adler, R.J. (1985). Extrema and level crossings of χ^2 processes. Report, Faculty of Industrial Engineering, Technion.

[6] Asmussen, S. (1984). Applied Probability. Institute of Mathematical Statistics, Copenhagen University: Book manuscript.

[7] Athreya, K.B. and Ney, P. (1978). A new approach to the limit theory of recurrent Markov chains. Trans. Amer. Math. Soc. 245, 493-501.[c] Berman, S.M. (1964). Limit theorems for the maximum term in stationary sequences. Ann. Math. Statist. 35, 502-516.

[8] Balkema, A.A., de Haan, L., and Resnick, S.I. (1984). Uniform rates of convergence to extreme value distributions. Statistical extremes and applications. Ed. by Tiago de Oliveira. Dordrecht: Reidel.

[9] Berman, S.M. (1962). Limiting distribution of the maximum term in sequences of dependent random variables. Ann. Math. Statist. 33, 894-908.

[10] Berman, S.M. (1964). Limit theorems for the maximum term in stationary sequences. Ann. Math. Statist. 35, 502-516.

[11] Berman, S.M. (1964). Limiting distribution of the maximum of a diffusion process. Ann. Math. Statist. 35, 319-329.

[12] Berman, S.M. (1982). Sojourns and extremes of stationary processes. Ann. Probab. 10, 1-46.

[13] Berman, S. (1983). High level sojourns of a diffusion process on a long interval. Z. Wahrsch. verw. Gebiete 62, 185-199.

[14] Berman, S.M. (1984). Limiting distribution of sums of nonnegative stationary random variables. Ann. Inst. Statist. Math. 36, 301-321.

[15] Bernstein, S. (1927) Sur l'extension du theoreme du calcul des probabilites aux sommes de quantites dependantes. Math. Ann. 97, 1-59.

[16] Billingsley, P. (1968). Convergence of probability measures. New York: Wiley.

[17] Blake, I.F. and Lindsey, W.C. (1973). Level-crossing problems for random processes. IEEE Trans. Inf. Theory 19, 295-315.

[18] Chernick, M.R. (1981). A limit theorem for the maximum of autoregressive processes with uniform marginal distributions. Ann. Probab. 9, 145, 149.

[19] Cohen, J.P. (1982a). The penultimate form of approximation to normal extremes. Adv. Appl. Prob., 14, 324-339.

[20] Cohen, J.P. (1982b). Convergence rates for the ultimate and penultimate approximations in extreme value theory. Adv. Appl. Prob., 14, 324-339.

[21] Cramer, H. (1965). A limit theorem for the maximum values of certain stochastic processes. Theory Probab. Appl. 10, 126-128.

[22] Darling, D.A. and Siegert, A.J. F. (1953). The first passage problem for a continuous Markov Process. Ann. Math. Statist. 24, 624-639.

[23] Davis, R. (1979). Maxima and minima of stationary sequences. Ann. Probab. 7, 453-460.

[24] Davis, R. (1982). Limit laws for the maximum and minimum of stationary sequences. Z. Wahrsch. verw. Gebiete 61, 31-42.

[25] Davis, R. (1982). Maximum and minimum of one-dimensional diffusions. Stochastic Process. Appl. 13, 1-9.

[26] Davis, R. (1982). The rate of convergence in distribution of the maxima. Statistica Neerlandica, 36, 31-35.

Davis, R. (1983). Limit laws for upper and lower extremes from stationary mixing sequences. J. Multivariate Anal. 13, 273-286.

Davis, R. (1983). Stable limits for partial sums of dependent random variables. Ann. Probab. 11, 262-269.

Davis, R. and Resnick, S.I. (1985). Limit theory for moving averages of random variables with regularly varying tail probabilities. Ann. Probab. 13, 179-

(1983). Point processes and multivariate extreme values. J. Mult. Anal. 13, 257-272.

[31] Durrett, R. and Resnick, S.I. (1978). Functional limit theorems for dependent variables. Ann. Probab. 6, 829-846.

[32] Falk, M. (1984). Uniform rates of convergence of extreme order statistics. Thesis, Department of Mathematics, University of Siegen.

[33] Finster, M. (1982). The maximum term and first passage time for autoregressions. Ann. Probab. 10, 737-744.

[34] Fisher, R.A. & Tippett, L.H.C. (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample. Proc. Camb. Phil. Soc. 24, 180-190.

[35] Frechet, M. (1927). Sur la loi de probabilite de l'ecart maximum. Ann. Soc. Math. Polon. 6, 93-116.

[36] Freedman, D. (1974). The Poisson approximation for dependent events. Ann. Probab., 256-269.

[37] Gnedenko, B.V. (1943). Sur la distribution limite du terme maximum d'une serie aleatoire. Ann. Math. 44, 423-453.

[38] Haan, L. de (1976). Sample extremes: an elementary introduction. Statist. Neerlandica 30, 161-172.

[39] Haan, L. de (1984). Slow variation and characterization of domains of attraction. Statistical extremes and applications. Ed. J. Tiago de Oliveira, Dordrecht: Reidel, 31-48.

[40] Haan, L. de and Hordijk, A. (1972). The rate of growth of sample maxima. Ann. Math. Statist. 43, 1185-1196.

[41] Haan, L. de and Resnick, S.I. (1977). Limit theory for multivariate sample extremes. Z. Wahrsch. verw. Gebiete 40, 317-337.

[42] Haan, L. de and Resnick, S.I. (1982). Local limit theorems for sample extremes. Ann. Probab. 10, 396-413.

[43] Haldane, J.B.S., and Jayakar, S.D. (1963). The distribution of extremal and nearly extremal values in samples from a normal population. Biometrika 50, 89-94.

[44] Hall, P. (1979). On the rate of convergence of normal extremes. J. Appl. Prob. 16, 433-439.

[45] Hall, P. (1980). Estimating probabilities for normal extremes. Adv. Appl. Prob. 12, 491-500.

[46] Hall, W.J. and Wellner, J.A. (1979). The rate of convergence in law of the maximum of an exponential sample. Statistica Neerlandica, 33, 151-154.

[47] Harter, H.L. (1978). A bibliography of extreme-value theory. Int. Statist. Rev. 46, 279-306.

[48] Hsing, T. (1984) Point processes associated with extreme value theory. Dissertation, Stoch. Proc. Center Tech. Rept. #83.

[49] Hsing, T. (1985). Extreme value theory for multivariate stationary sequences. Report, Department of Mathematics, University of Texas at Arlington.

[50] Hsing, T., Husler, J., and Leadbetter, M.R. (1985) On the exceedance point process for a stationary sequence. To appear in Probability Theory and Related Fields.

[51] Ito, K. (1969). Stochastic processes. Lecture notes series #16, Department of Mathematics, Aarhus University.

[52] Ito, K. and McKean, H.P. (1965). Diffusion processes and their sample paths. Berlin: Springer.

[53] Kallenberg, O. (1976) Random measures. Academic Press London, New York.

[54] Leadbetter, M.R. (1983) Extremes and local dependence in stationary sequences. Zeit. fur wahrsch. verw. Geb. $\underline{65}$, 291-306.

[55] Leadbetter, M.R., Lindgren, G., Rootzen, H. (1983). Extremes and related properties of random sequences and processes. Springer Series in Statistics. Springer-Verlag N.Y.

[56] LePage, R., Woodroofe, M., and Zinn, J. (1981). Convergence to stable distributions via order statistics. Ann. Probab. 9, 624-632.

[57] Lindgren, G. (1980). Extreme values and crossings for the χ^2-process and other functions of multidimensional Gaussian processes, with reliability applications Adv. Appl. Probab. 12, 746-774.

[58] Lindgren, G. (1980). Point processes of exits by bivariate Gaussian processes and extremal theory for the χ^2-process and its concomitants. J. Multivariate Anal. 10, 181-206.

[59] Lindgren, G. (1984). Extremal ranks and transformation of variables for extremes of functions of multivariate Gaussian processes. Stochastic Process. Appl. 17, 285-312.

[60] Lindgren, G. (1984). Use and structure of Slepian model processes for prediction and detection in crossing and extreme value theory. In: Statistical extremes and applications. Ed. Tiago de Oliveira, Dodrecht: Reidel.

[61] Loynes, R.M. (1965) Extreme values in uniformly mixing stationary stochastic processes. Ann. Math. Statist. $\underline{36}$, 993-999.

[62] Mandl, P. (1968). Analytic treatment of one-dimensional Markov processes. New York: Springer.

[63] Mises, R. von (1936). La distribution de la plus grande de n valeurs. Reprinted in: Selected papers II. Amer. Math. Soc., Providence, R.I. 1954, 271-294.

[64] Mittal, Y. and Ylvisaker, D. Limit distributions for the maxima of stationary Gaussian processes Stoch. Proc. Applns. 3, 1-18.

[65] Mori, T. (1977) Limit distributions of two-dimensional point processes generated by strong mixing sequences Yokohama Math. J. 25, 155-168.

[66] Newell, G.F. (1962). Asymptotic extreme value distributions for one-dimensional diffusion processes. J. Math. Mech. 11, 481-496.

[67] Nummelin, E. (1978). A splitting technique for Harris recurrent Markov chains. Z. Wahrsch. verw. Gebiete 43, 309-318.

[68] O'Brien, G. (1984). Extreme values for stationary and Markov sequences. Report, Department of Mathematics, York University.

[69] Pickands, J. III (1969). Asymptotic properties of the maximum in a stationary Gaussian process. Trans. Amer. Math. Soc. 145, 75-86.

[70] Pickands, J. III (1969). Upcrossing probabilities for stationary Gaussian processes. Trans. Amer. Math. Soc. 145, 51-73.

[71] Pickands, J. III (1971). The two-dimensional Poisson Process and extremal processes. J. Appl. Probab. 8, 745-756.

[72] Pickands, J. III (1980). Multivariate negative exponential and extreme value distributions. Manuscript, Wharton School of Business, University of Pennsylvania.

[73] Resnick, S.I. (1985). Uniform rates of convergence to extreme value distributions. Report, Department of Statistics, Colorado State University.

[74] Resnick, S.I. (1986). Point processes, Regular variation and weak convergence. Adv. Appl. Probab., 18, 66-138.

[75] Rootzen, H. (1978). Extremes of moving averages of stable processes. Ann. Probab. 6, 847-869.

[76] Rootzen, H. (1983). Extreme value theory for moving average processes. Ann. Probab. 14, 612-652.

[77] Rootzen, H. (1983). The rate of convergence of extremes of stationary normal sequences. Adv. Appl. Probab. 15, 54-80.

[78] Rootzen, H. (1985). A ratio limit theorem for weighted sums. Report, Department of Mathematical Statistics, University of Copenhagen.

[79] Rootzen, H. (1985). Maxima and exceedances of stationary Markov chains. Adv. Appl. Probabl., to appear.

[80] Sharpe, K. (1978). Some properties of the crossings process generated by a stationary χ^2 process. Adv. Appl. Probab. 10, 373-391.

[81] Slepian, C. (1962) The one-sided barrier problem for Gaussian noise Bell

Syst. Tech. J. $\underline{41}$, 463-501.

[82] Smith, R. (1982). Uniform rates of convergence in extreme value thoery. Adv. Appl. Prob., $\underline{14}$, 543-565.

[83] Tiago de Oliveira, J., Bibliography of extreme value theory, to appear.

[84] Volkonski and Rozanov (1959) Some limit theorems for random functions I. Theor. Probab. Appl. $\underline{4}$, 178-197.

[85] Watts, V., Rootzen, H. Leadbetter, M.R. (1982). Convergence in distribution of intermediate order statistics from stationary sequences. Ann. Probab. $\underline{10}$, 653-662.

[86] Weissman, I. (1984). On the asymptotic joint distribution of an unbounded number of sample extremes. Report, Faculty of Industrial Engineering, Technion.

EVD

5— 82

DTIC